Complexity

Cognitive Network Science: A New Frontier

Lead Guest Editor: Yoed Kenett
Guest Editors: Nicole Beckage, Cynthia Siew, and Dirk Wulff

Cognitive Network Science: A New Frontier

Complexity

Cognitive Network Science: A New Frontier

Lead Guest Editor: Yoed Kenett
Guest Editors: Nicole Beckage, Cynthia Siew, and Dirk Wulff

Contents

Cognitive Network Science: A New Frontier
Yoed N. Kenett, Nicole M. Beckage, Cynthia S. Q. Siew, and Dirk U. Wulff
Editorial (4 pages), Article ID 6870278, Volume 2020 (2020)

From Topic Networks to Distributed Cognitive Maps: Zipfian Topic Universes in the Area of Volunteered Geographic Information
Alexander Mehler, Rüdiger Gleim, Regina Gaitsch, Wahed Hemati, and Tolga Uslu
Research Article (47 pages), Article ID 4607025, Volume 2020 (2020)

Network Growth Modeling to Capture Individual Lexical Learning
Nicole M. Beckage and Eliana Colunga
Research Article (17 pages), Article ID 7690869, Volume 2019 (2019)

Cognitive Network Science: A Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics
Cynthia S. Q. Siew, Dirk U. Wulff, Nicole M. Beckage, and Yoed N. Kenett
Review Article (24 pages), Article ID 2108423, Volume 2019 (2019)

Analyzing Knowledge Retrieval Impairments Associated with Alzheimer's Disease Using Network Analyses
Jeffrey C. Zemla and Joseph L. Austerweil
Research Article (12 pages), Article ID 4203158, Volume 2019 (2019)

Mediation Centrality in Adversarial Policy Networks
Stefan M. Herzog and Thomas T. Hills
Research Article (15 pages), Article ID 1918504, Volume 2019 (2019)

Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation
Karl D. Neergaard and Chu-Ren Huang
Research Article (21 pages), Article ID 6979830, Volume 2019 (2019)

Expanding Network Analysis Tools in Psychological Networks: Minimal Spanning Trees, Participation Coefficients, and Motif Analysis Applied to a Network of 26 Psychological Attributes
Srebrenka Letina, Tessa F. Blanken, Marie K. Deserno, and Denny Borsboom
Research Article (27 pages), Article ID 9424605, Volume 2019 (2019)

Human Sensitivity to Community Structure Is Robust to Topological Variation
Elisabeth A. Karuza, Ari E. Kahn, and Danielle S. Bassett
Research Article (8 pages), Article ID 8379321, Volume 2019 (2019)

The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning
R. Harald Baayen, Yu-Ying Chuang, Elnaz Shafaei-Bajestan, and James P. Blevins
Research Article (39 pages), Article ID 4895891, Volume 2019 (2019)

Spread the Joy: How High and Low Bias for Happy Facial Emotions Translate into Different Daily Life Affect Dynamics
Charlotte Vrijen, Catharina A. Hartman, Eeske van Roekel, Peter de Jonge, and Albertine J. Oldehinkel
Research Article (15 pages), Article ID 2674523, Volume 2018 (2018)

Cohort and Rhyme Priming Emerge from the Multiplex Network Structure of the Mental Lexicon
Massimo Stella
Research Article (14 pages), Article ID 6438702, Volume 2018 (2018)

Hindawi
Complexity
Volume 2020, Article ID 6870278, 4 pages
https://doi.org/10.1155/2020/6870278

Editorial

Cognitive Network Science: A New Frontier

Yoed N. Kenett[1], **Nicole M. Beckage**[2], **Cynthia S. Q. Siew**[3] and **Dirk U. Wulff**[4,5]

[1]*Department of Psychology, University of Pennsylvania, Philadelphia, USA*
[2]*Intel Labs, Hillsboro, USA*
[3]*National University of Singapore, Singapore*
[4]*Center for Cognitive and Decision Science, University of Basel, Basel, Switzerland*
[5]*Max Planck Institute for Human Development, Berlin, Germany*

Correspondence should be addressed to Yoed N. Kenett; yoedk@sas.upenn.edu

Received 8 January 2020; Accepted 9 January 2020; Published 28 April 2020

A major challenge in studying the complexity of cognition relates to quantifying abstract theoretical cognitive constructs, such as language, memory, or thinking, and studing the representation of these abstract constructs. Such quantifications of these abstract constructs are based on indirect measures of cognitive systems such as behavioral measures or neural activity. In the past two decades, an increasing number of studies have used network science methods to study complex systems.

Network science is based on mathematical graph theory and offers quantitative methods to investigate complex systems [1]. A network is made up of nodes, which represent the basic unit of the system (e.g., concepts in semantic memory) and links, or edges, which signify the relations between them (e.g., semantic similarity). While the application of network science methodologies has become an extremely popular approach to study brain structure and function [2], it has been used to study cognitive phenomena to a much lesser extent, despite classic cognitive theory in language and memory being highly related to a network perspective [3].

So far, the application of network science to cognitive science has enabled the direct examination of the theory that highly creative individuals have a more flexible semantic memory structure [4], identified mechanisms of language development through network growth modeling [5], shed novel light on statistical learning [6], examined phonological and orthographic effects [7, 8], provided new insight into the structure of semantic network of second language in bilinguals [9], and studied changes in memory structure across the lifespan [10].

The aim of this special issue is to demonstrate the potential and strength of applying network science methods to study cognition (broadly defined). In the article "Cognitive Network Science: A Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics," C. S. Q. Siew et al. provide a comprehensive review on the field of Cognitive Network Science. Specifically, their article is focused on three key main theses: (1) Network science provides a quantitative approach to represent cognitive systems; (2) network science facilitates a deeper understanding of human cognition by allowing the researcher to consider how network structure and the processes operating on the network structure interact to produce behavioral phenomena; and (3) network science provides a framework to model structural changes in cognitive systems at multiple scales.

It is striking that, without any prior arrangement, the collection of articles in this special issue has aligned rather well with the main theses of the comprehensive review by C. S. Q. Siew et al. Articles by M. Stella, K. D. Neergaard, and C.-R. Huang, S. Letina et al., R. H. Baayen et al., S. M. Herzog and T. T. Hills, A. Mehler et al., and C. Vrijen et al. illustrate how network science methods can be used to *represent* a variety of cognitive, linguistic, psychological, and even social systems. Articles by M. Stella, K. D. Neergaard and C.-R. Huang, R. H. Baayen et al., S. M. Herzog and T. T. Hills, C. Vrijen et al., and E. A. Karuza et al. demonstrate how the structure of the cognitive network plays an important role in predicting *behavioral outcomes* in domains including language comprehension and production, statistical learning,

mental health, and conflict resolution. Finally, articles by N. M. Beckage and E. Colunga and J. C. Zemla and J. L. Austerweil focus on modeling *structural changes* in the network representation as children learn new words and as cognitive decline sets in.

Furthermore, as apparent from the article summaries below, the collection of articles in this special issue shows how network science approaches can be flexibly applied to address a broad range of topics and domains in the cognitive and social sciences, as well as how network science approaches can creatively advance methodology in these areas. Articles by M. Stella, K. D. Neergaard and C.-R. Huang, A. Mehler et al., and R. H. Baayen et al. show how various aspects of the *mental lexicon* can be represented as a cognitive network. Articles by E. A. Karuza et al. and N. M. Beckage and E. Colunga focus on how humans *learn* temporal, event-based visual information, and language, respectively. Articles by S. Letina et al. and C. Vrijen et al. analyzed *psychological* networks of personality attributes and affect dynamics. Finally, other articles focused on the *social* graphs of mediators (S. M. Herzog and T. T. Hills), modeling of *cognitive decline* (J. C. Zemla and J. L. Austerweil), and network *methodology* (A. Mehler et al. and S. Letina et al.).

Taken together, the articles in this special issue demonstrate the feasibility, and strength, of applying the quantitative language of network science to advance our understanding of complex cognitive phenomena. We present a brief overview of each of the articles in this special issue, according to the order in which they were published.

In the article "Cohort and Rhyme Priming Emerge from the Multiplex Network Structure of the Mental Lexicon," M. Stella used a multiplex lexical network representing both semantic and phonological relationships among words in the mental lexicon to examine two aspects of phonological priming: cohort priming and rhyme priming. Results indicated that both cohort words (i.e., words that share the same initial sounds) and rhyme words (i.e., words that rhyme) were "closer" in terms of distance computed on various layers in the multiplex as compared to random expectation. These results suggest an alternative account of priming effects in psycholinguistics, whereby facilitatory priming may simply emerge as a consequence of higher-order structural relationships among words.

In the article "Spread the Joy: How High and Low Bias for Happy Facial Emotions Translate into Different Daily Life Affect Dynamics," C. Vrijen et al. examined how daily life affect dynamics differed among individuals with low and high levels of bias toward happy facial emotions. Daily-life affect networks refer to networks that represent different emotions (positive/negative) and the effect from one time interval of six hours to the next on these emotions. Specifically, the aim of this study was to examine the importance of laboratory measurement of happy bias in peoples' daily life. Combining a network psychometric approach with experience sampling methodology, the authors found that individuals with high happy bias showed more sustained effects of positive, rewarding experiences in their affect networks over time as compared to individuals with low happy bias. These results suggest that sensitivity to positive experiences may be related to a bias for happy emotions and may act as a buffer against the development of depression.

In the article "The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded not in (De) Composition but in Linear Discriminative Learning," R. H. Baayen et al. rely on a neural network, trained sentence-by-sentence, to predict the occurrence of lexemes. This trained model predicts a variety of behaviors including paired associate learning and semantic relatedness ratings. Moreover, when combined with a phonological representation that maps phonemes to words, they found that the resulting representation allowed them to account for the behaviors recruiting the entire pipeline of visual and auditory comprehension, from word form to meaning. The article suggests methods and applications for learned network representations and how those representations may offer cognitive insight and predict behavior in linguistic experiments.

In the article "Human Sensitivity to Community Structure Is Robust to Topological Variation", E. A. Karuza et al. replicate and extend their previous work by examining how the topology of the environment facilitates statistical learning. E. A. Karuza et al. previous work showed learner sensitivity to the presence of community structure within temporal sequences. However, whether such a sensitivity generalizes to variations in graph topology was unknown. To address this, the authors systematically vary the number and size of communities and assess how it impacts learning. The authors show that learners are sensitive to community structure across a range of network topologies (that vary in their number and size of communities). Thus, this work demonstrates how network science methods can be used to study how individuals are sensitive to the topology of their environment.

In the article "Expanding Network Analysis Tools to Psychological Networks: Minimal Spanning Trees, Participation Coefficients, and Motif Analysis Applied to a Network of 26 Psychological Attributes," S. Letina et al. turn to minimum spanning trees and motif analysis to studying the emerging hierarchy of psychological trait networks. The authors derive a network based on a variety of psychological concepts (correlations of self-reported personality traits from questionnaires such as the Schwatz Value Survey, the Big Five personality traits, Sensational Interest Questionnaire, and others). From this weighted network, they define a minimum spanning tree, participation coefficient, and observed motifs in the original network to study the relationship between these measures and psychological constructs. The authors show how these three types of network analysis, not currently used to study psychological constructs, provides meaningful information and complement each other in the ability to capture and explain the interaction of psychological traits. The authors conclude that certain traits, such as empathy, are central to the network and other nodes, such as intelligence, which are in the periphery still are related to a large number of other traits.

In the article "Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation", K. D. Neergaard and C.-R. Huang used network science methods to construct various types of

phonological networks of Mandarin Chinese. These phonological networks were constructed based on various phonological annotation strategies, inferred from a Chinese phonological association task. In this phonological association task, participants produced Chinese syllables that sounded similar to a target Chinese syllable. The authors then use RT data from the phonological association task to identify the optimal annotation strategy to construct the Chinese phonological network. The results indicated that structural aspects of the Chinese phonological network influenced how people "search" for similar sound neighbors in the phonological lexicon. Thus, the authors present a method to systematically study phonological segmentation of languages and how network science can be used to examine how the structure of such an optimally segmented phonological system influence "search processes" operating over it.

In the article "Mediation Centrality in Adversarial Policy Networks," S. M. Herzog and T. T. Hills introduce and explore a new network measure - mediation centrality, a network measure for identifying mediators in bipartite adversarial networks. Adversarial systems can be defined as systems composed of individuals with opposing views, such as Democrats versus Republicans in US politics. Adversarial networks can be represented by bipartite networks, where individuals are connected by edges to the views they support. Over such a bipartite network, a good mediator is an individual (node) that can minimize the polarity of such opposing views. Thus, mediation centrality is computed by combining centrality metrics from subgraph projections where the projections are defined in relation to different sets of views. The authors argue that this measure is important in identifying mediators who can advance conflict resolution in polarized adversarial systems. Finally, S. M. Herzog and T. T. Hills demonstrate the utility of computing mediation centrality across a range of examples, demonstrating its fruitfulness in adversarial systems.

In the article "Analyzing Knowledge Retrieval Impairments Associated with Alzheimer's Disease Using Network Analysis," J. C. Zemla and J. L. Austerweil employ a sophisticated, Bayesian approach to infer an individual's semantic network from just a few number of verbal fluency sequences. The approach is elegant as it is based on a complete cognitive model, encompassing a search process retrieving from an underlying, to-be-inferred representation. Using their modeling approach, J. C. Zemla and J. L. Austerweil are able to generate novel, actionable insights concerning the cognitive development of patients with Alzheimer's disease. Specifically, they show the semantic networks of patients with Alzheimer's disease are less connected, more disordered, and, generally, less small-world-like than those of healthy controls.

In the article "Network Growth Modeling to Capture Individual Lexical Learning," N. M. Beckage and E. Colunga introduce a network growth modeling framework for quantifying the influence of (1) different network representations, (2) growth processes, and (3) node importance. They test their network growth framework on the prediction of individual language learning trajectories. Their models provide quantification on the emergent structure of young toddler's vocabularies and provide a set of tools to study individual differences in language acquisition trajectories. They show evidence that the acquisition model is influenced by the underlying network representation, the assumed growth process, and the network centrality measure used to quantify the importance of words, highlighting the complex and multifaceted nature of early acquisition. Their framework also provides new tools of analysis and suggests new hypotheses that can be tested in experimental interventions in language development and are targeted at the level of the individual child's current knowledge.

In the article "From Topic Networks to Distributed Cognitive Maps Zipfian Topic Universes in the Area of Volunteered Geographic Information," A. Mehler et al. introduce a set of novel methods to extend the standard analysis of co-occurrence networks from textual corpora based on a multiplex network approach. Specifically, the authors define an approach that allows for thematic comparison directly between different communities by deriving a network of topics from varied sources of information, such as different readership, different authorship, and different medium. The resulting framework introduces a process for deriving such network layers as (1) author topic networks in which connected authors tend to refer to similar thematic elements throughout their writing, (2) text networks which capture the relationship of a single document with other text documents, (3) constituent layers which can be defined to capture such relationships as lexicographic and phrasal information, and (4) contextual layers which link topics based on such high-level features as media and genre. This multiplex topic network approach could allow for modeling of social and cognitive interactions from text-based information sources such as those found on the world wide web.

Conflicts of Interest

Yoed Kenett declares that he has worked in the past with Dr. Elisabeth Karuza. Nicole Beckage and Yoed Kenett declare that they have worked in the past with Dr. Joseph Austerweil. Dirk Wulff, Cynthia Siew, and Nicole Beckage declare that they have worked in the past with Dr. Thomas Hills. Each editor did not handle work by the named individual they have previously worked with.

Yoed N. Kenett
Nicole M. Beckage
Cynthia S. Q. Siew
Dirk U. Wulff

References

[1] A. Baronchelli, R. Ferrer-i-Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen, "Networks in cognitive science," *Trends in Cognitive Sciences*, vol. 17, no. 7, pp. 348–360, 2013.

[2] J. D. Medaglia, M.-E. Lynall, and D. S. Bassett, "Cognitive network neuroscience," *Journal of Cognitive Neuroscience*, vol. 27, no. 8, pp. 1471–1491, 2015.

[3] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing," *Psychological Review*, vol. 82, no. 6, pp. 407–428, 1975.
[4] Y. N. Kenett and M. Faust, "A semantic network cartography of the creative mind," *Trends in Cognitive Sciences*, vol. 23, no. 4, pp. 271–274, 2019.
[5] N. Beckage, L. Smith, and T. T. Hills, "Small worlds and semantic network growth in typical and late talkers," *PLoS One*, vol. 6, no. 5, Article ID e19348, 2011.
[6] E. A. Karuza, S. L. Thompson-Schill, and D. S. Bassett, "Local patterns to global architectures: influences of network topology on human learning," *Trends in Cognitive Sciences*, vol. 20, no. 8, pp. 629–640, 2016.
[7] C. S. Q. Siew, "Community structure in the phonological network," *Frontiers in Psychology*, vol. 4, p. 553, 2013.
[8] C. S. Q. Siew, "The orthographic similarity structure of English words: insights from network science," *Applied Network Science*, vol. 3, no. 1, p. 13, 2018.
[9] K. Borodkin, Y. N. Kenett, M. Faust, and N. Mashal, "When pumpkin is closer to onion than to squash: the structure of the second language lexicon," *Cognition*, vol. 156, pp. 60–70, 2016.
[10] D. U. Wulff, S. De Deyne, M. N. Jones, R. Mata, and T. A. L. Consortium, "New perspectives on the aging lexicon," *Trends in Cognitive Sciences*, vol. 23, no. 8, pp. 686–698, 2019.

Hindawi
Complexity
Volume 2020, Article ID 4607025, 47 pages
https://doi.org/10.1155/2020/4607025

WILEY | Hindawi

Research Article

From Topic Networks to Distributed Cognitive Maps: Zipfian Topic Universes in the Area of Volunteered Geographic Information

Alexander Mehler,[1] **Rüdiger Gleim**,[1] **Regina Gaitsch**,[2] **Wahed Hemati**,[1] **and Tolga Uslu**[1]

[1]*Goethe-University Frankfurt, Frankfurt, Germany*
[2]*Justus Liebig University Giessen, Giessen, Germany*

Correspondence should be addressed to Alexander Mehler; mehler@em.uni-frankfurt.de

Received 30 August 2018; Revised 7 June 2019; Accepted 2 July 2019; Published 27 April 2020

Guest Editor: Nicole Beckage

Are nearby places (e.g., cities) described by related words? In this article, we transfer this research question in the field of lexical encoding of geographic information onto the level of intertextuality. To this end, we explore *Volunteered Geographic Information* (VGI) to model texts addressing places at the level of cities or regions with the help of so-called topic networks. This is done to examine how language encodes and networks geographic information on the aboutness level of texts. Our hypothesis is that the networked thematizations of places are similar, regardless of their distances and the underlying communities of authors. To investigate this, we introduce *Multiplex Topic Networks* (MTN), which we automatically derive from *Linguistic Multilayer Networks* (LMN) as a novel model, especially of thematic networking in text corpora. Our study shows a Zipfian organization of the thematic universe in which geographical places (especially cities) are located in online communication. We interpret this finding in the context of *cognitive maps*, a notion which we extend by so-called *thematic maps*. According to our interpretation of this finding, the organization of thematic maps as part of cognitive maps results from a tendency of authors to generate shareable content that ensures the continued existence of the underlying media. We test our hypothesis by example of special wikis and extracts of Wikipedia. In this way, we come to the conclusion that geographical places, whether close to each other or not, are located in neighboring semantic places that span similar subnetworks in the topic universe.

1. Introduction

In this article, we explore crowd-sourced resources for automatically characterizing geographical places with the help of so-called *topic networks*. Our goal is to model the thematic structure of corpora of natural language texts that are about certain places seen as thematic frames. This is done in order to automatically compare the thematic structures of corpora of texts about these places, which will be represented as topic networks. In this way, we want to investigate the regularity or systematicity according to which geographical objects (i.e., cities and regions) are dealt with, especially in online communication.

Our work relates to what is described by Crooks et al. [1] as a novel paradigm of modeling "*urban morphologies*." We not only add special wikis such as regional and city wikis as candidates to the resources listed in [1] but also introduce a novel method for modeling their content. This concerns local media of collaborative writing about places (cf. [2]), which contain *everyday place descriptions* [3] authored and networked according to the wiki principle. The corresponding wikis and the subgraphs of Wikipedia that we additionally analyze manifest *Volunteered Geographic Information* (VGI) [4–6] and thus relate to what is called the wikification of *Geographical Information Systems* (GIS) [7]. VGI is "*completing traditional authoritative geographic information*" [8], an information source which is still "*underutilized*" in geography [9] as a source of big textual data [8] making natural language processing an indispensable prerequisite for its analysis. According to Hardy

et al. [6], authoring VGI has a *spatial* component in the sense that people likely write about *local* content though this also holds for Wikipedia for a minor degree [10]. This spatial component can be accompanied by a lack of quality assurance, which makes VGI susceptible to deficiencies and to a distorted resource of still unknown extent [5]. In any event, the biased coverage of VGI is a characteristic of resources like Wikipedia so that the same region can be displayed very differently in its various language editions [11], a sort of biasing which is typical for user-generated content. Nevertheless, Hahmann and Burghardt [12] show that more than 50% of the articles in the German Wikipedia contain georeferenced data (at least indirectly via links to other articles), so that such media can be regarded as rich resources of VGI. Moreover, Goodchild and Li [5] point to the fact that crowd-sourcing or, more precisely, crowd-curation [13], as enabled by wikis, is a means of quality assurance.

We follow this concept and assume that geographic data, as manifested linguistically in online media, are a valuable resource to investigate how communities form a common sense for addressing places of common interest. In line with Clare ([14], 41), we additionally assume that "*[a]s people communicate more about a place, social consensus will create increased similarity between and within people's judgments of it*." However, we also assume that the latter similarity can affect communications of different communities about different places. In this way, we assume a kind of horizontal self-similarity [15] of the thematic structure of online media, which is more or less independent of the underlying theme and the community. That is, our hypothesis on the theming of places is as follows.

Hypothesis 1. Thematizations of different places at a certain level of thematic abstraction tend to be similar among each other (rather than being dissimilar) (1) in the sense that they focus on similar topics and (2) the way these topics are networked and (3) with respect to the skewness of this focus, regardless of whether the underlying media are generated by different communities and whether these communities address related or unrelated places at near or distant spaces.

The intuition behind Hypothesis 1 is that thematizations of places in web-based communication are seemingly somehow thematically redundant: in reporting, for example, on the cities in which people live, they may aim to emphasize the special character of these places. It seems, however, as if a thematic trend is breaking ground that ultimately makes such reports appear thematically very similar. Whether or not this intuition is actually a trend that can be observed specifically in the field of wiki-based media is something this study is intended to clarify. From this point of view, it is obvious that Hypothesis 1 is only a starting point which in itself needs further clarification in order to be testable: similarity, for example, is a highly context-sensitive attribute [17] that needs further definitional specifications in order to be computable. Likewise, the concept of thematization (theme or topic)—a concept which according to Adamzik [18] has so far found comparatively less attention in linguistics—is not yet specified in Hypothesis 1. Thus, an appropriate elaboration and concretization of Hypothesis 1 is one of the main tasks of the present paper. To this end, it is developing a *generic topic network model* in conjunction with a measurement procedure which will specify both the notion of *similarity* (which will be defined in terms of the graph similarity of topic networks) and of the *thematization of places* (which will be defined in terms of topic labeling and topic networking). This topic network model will allow Hypothesis 1 to be reformulated and concretized in the form of variants (i.e., Hypotheses 2–4), which will be presented in Section 3.2.7 and whose formulations presuppose the topic network model that this paper develops in the preceding sections.

The skewness that is mentioned by Hypothesis 1 reminds one of a Zipfian process, according to which a few topics dominate, while the majority of candidate topics are underrepresented or disregarded. Therefore, we speak of *Zipfian thematic universes*, which are spanned by the thematization of the same places in online media such as special wikis of the sort studied here. By the term *topic*, we refer to the notion of aboutness of texts [18, 19]. From a linguistic point of view, the terminology of Hypothesis 1 seems to be confusing when referring to places as *what is given* and with topic to *what is said about these places*. The reason is that linguistics distinguishes between what is given (*theme* or *topic*) and what is said about it (*rheme*, *comment*, or *focus*) in a given piece of text [18, 20–22]: a mention of a city like *Vienna*, for example, can be connected with certain subtopics (e.g., *classical music*), which characterize this place rhematically by providing new information about it. The latter distinction is meant when we relate subtopics in the role of rhemes to places in the role of topics in the linguistic sense. Thus, when talking about topics as part of a computational model, we will use the term *topic* (topic_2), while when talking about places as topics in the linguistic sense (topic_1), we will use the term *theme* and speak about its rhemes as its subtopics modeled by topics (topic_2) as units of our model. This scenario and its relation to Hypothesis 1 are depicted in Figure 1. It shows a generalization of a hypothesis of Louwerse and Zwaan [16] according to which language encodes geographical information: the places p and q, which are understood as conceptual units (i.e., mental models), are described by or expressed in two discourse units (texts, dialogs, etc.) x and y. From the latter units, the topic representations α and β are derived by means of a computational model (e.g., *Latent Dirichlet Allocation* (LDA) [23] or the topic network model introduced in Section 3). While such derived topics are part of the computational model, the underlying discourses belong to the modeled system. We assume that the conceptual unit p (q) is structured into a system of networked rhemes or subtopics p_i (q_m). Ideally, the derived topic α in Figure 1 is a valid model of one of the rhemes of place p (e.g., p_i) and β of one of the rhemes of place q (e.g., q_m). If we assume now that p and q are conceptually related (e.g., similar) to each other, then the linguistic encoding hypothesis implies that this is possibly reflected by a relatedness (e.g., similarity) relation among some rhemes of these places (e.g., by the relatedness of p_i and q_m). From the point of view of modeling, this relation is ideally mapped by the relatedness (e.g., similarity)

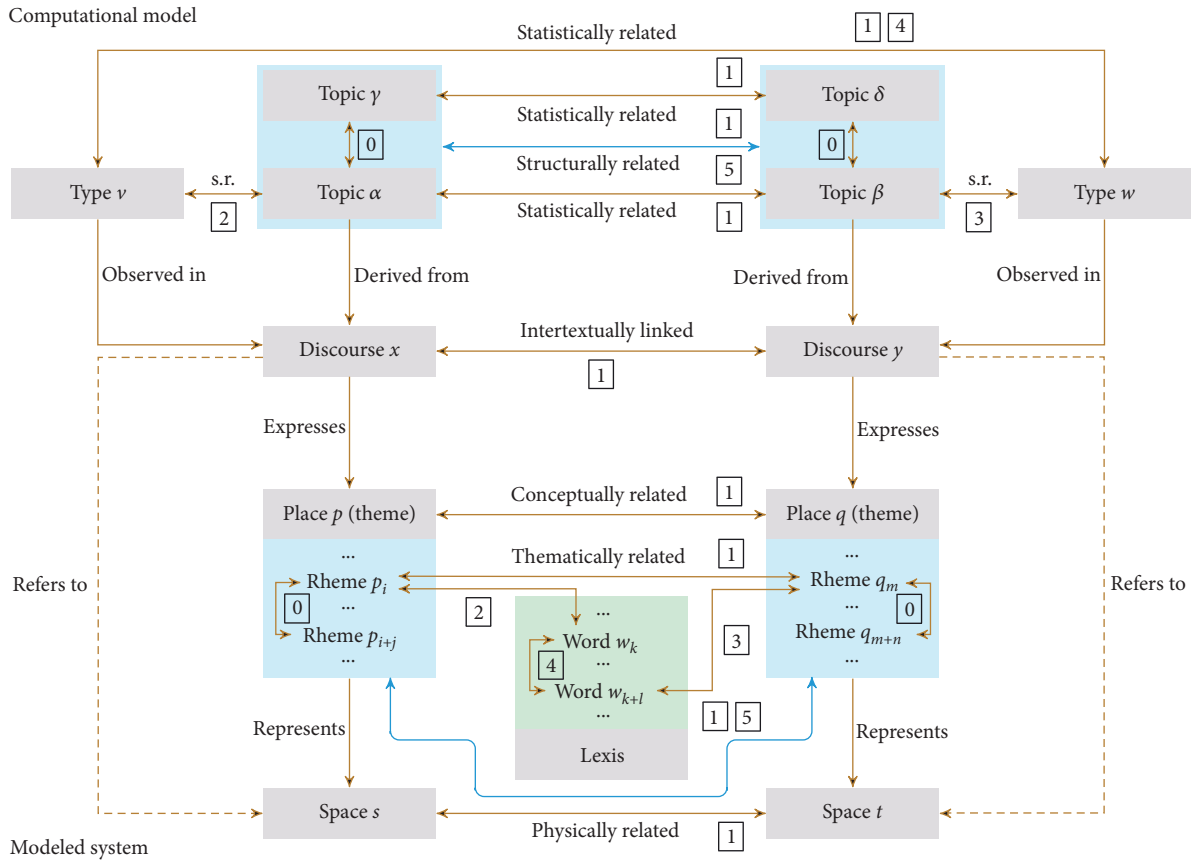


FIGURE 1: Schematic depiction of a generalization of a hypothesis of Louwerse and Zwaan [16] saying that language encodes geographical information: the places p and q are expressed in the discourses x and y, respectively, from which the topic representations α and β are computationally derived. Places are structured into systems of networked rhemes or subtopics. The conceptual relatedness of p and q is grounded in the relatedness of the rhemes p_i and q_m and modeled by the relatedness of the derived topics α and β modeling these rhemes. According to the semiotic triangle, we assume that the relation of signs (here, texts) to their referents (here, spaces) is mediated by sign processes. We use dashed arcs to express the indirect relation of the former to the latter. In lexical variants of this approach, p and q are preferably denoted or described by some words w_k and w_{k+l} of the underlying lexis, which are syntagmatically or paradigmatically associated and modeled by some types v and w. Framed numbers indicate relations that potentially parallelize each other. *s.r.* means *statistically related*.

of the derived topics α and β. We assume that conceptual relations between places can be parallelized by relations of physical proximity or distance between spaces that are mentally modeled by these places. If one additionally assumes that proximity in space correlates with relatedness in conceptual space (the less the distant, the more the similar, for example), one obtains a linguistic variant of Tobler's so-called first law (see Section 2). If we look at the literature (see Section 2), we find that the approaches in this area differ in terms of the linguistic level at which they observe the linguistic encoding of platial [13] relations: for example, at the level of intertextually linked texts, at the level of the topics these texts are about, or at the level of lexical elements used by these and other texts to deal with the latter topics. In lexical variants of this approach, the places p and q, for which we assume that they are conceptually related, are preferably referred to or described by means of lexical items w_k and w_{k+l} (see Figure 1) of the underlying lexis that are syntagmatically or paradigmatically associated. From the point of view of modeling, we have to then assume the two types v and w (as models of the words w_k and w_{k+l}) for which we automatically detect, for example, their (paradigmatic) closeness in semantic space (cf. [24, 25]) or the similarity of their (syntagmatic) co-occurrence statistics (cf. [26]).

From this analysis, we obtain a series of reference points or means for encoding geographical information about conceptual relations (see [1] in Figure 1) of places. This concerns more precisely a series of possible parallelizations of such relations, which may ultimately be parallelized by relations between the spaces designated by these places (for the numbers in brackets, see Figure 1): at the level of the modeled system, this refers to thematically linked rhemes, intertextually linked discourse units (e.g., texts), and syntagmatically or paradigmatically linked words ([1]). From a modeling point of view, we distinguish the statistical relatedness of types or of topics as candidate parallelizations ([1]). Beyond that, we find the parallelization of the relatedness of rhemes and words on the one hand and of types and topics on the other ([2], [3]), as well as that of the relatedness of words on the one hand and of types on the other ([4]). The parallelization of the relatedness of rhemes of the same place ([0]) by the relatedness of the rhemes of

another place concerns the core of our network approach. Such relations among rhemes constitute rhematic networks or networks of rhemes on both sides of the affected places. Our main assumption is now that any such rhematic network, which manifests the thematic structure of a place, can be related *as a whole* to that of another place. In doing so, it is, from a modeling point of view, ideally parallelized *by the structural relatedness (e.g., similarity or complementarity) of topic networks*, which are derived from corpora of texts, each of which describes one of these places ([5]). This type of parallelization affects entire networks of linguistic objects and yet offers a means of encoding the conceptual relationship of places ([1]) or the proximity of spaces, respectively. In the present paper, we explore relations of Type [5] in order to learn about the encoding of geographical information in natural language texts, that is, about relations of Type [1]. To this end, we develop, instantiate, and empirically test a formal model of multiplex topic networks derived from so-called linguistic multilayer networks as a model of relations of Type [5].

From this point of view, Hypothesis 1 means that certain rhemes of places and the structure they span resemble each other, regardless of how far the quantified distances of the spaces represented by these places are and regardless of the fact that the texts in which these rhemes are described are written by different communities. To test this hypothesis, we introduce topic networks to make the networking of topics a research object according to the scenario described in Figure 1, that is, in relation to the hypothesis of linguistic encoding of geographical information. The contributions of this article are of theoretical, methodical, and empirical nature.

(1) Formal modeling: we develop a generic, extensible formalism for the representation of topic networks that cover a wide range of informational sources for spanning and weighting topic links. To this end, we introduce the notion of *multiplex topic networks* derived from so-called *multilayer linguistic networks*. In this way, we enable the same place to be represented by a family of thematic networks that offer different perspectives on the networking of its rhemes. We exemplify this model by means of two perspectives provided by so-called *Text Topic Networks* (TTN) and their corresponding *Author Topic Networks* (ATN).

(2) Procedural modeling: we develop a measurement procedure for instantiating our formal model. To this end, we introduce novel measures of the similarity of labeled graphs that are sensitive to their links and to their nodes.

(3) Experimentation: we further develop the range of baseline statistics in network theory in order to better assess the quality of our measurements. To this end, we test our model by means of a threefold classification experiment that compares a set of TTNs with each other, a set of corresponding ATNs with each other, and the former TTNs with the latter ATNs.

(4) Theory formation: we interpret our findings in the context of cognitive maps, thus building a bridge between our network-theoretical approach and approaches to the cognitive representation of geographical information. We show how to integrate the analysis of entire networks into the research about the linguistic encoding of geographical information (see Figure 1).

This paper is organized as follows: Section 2 discusses related work. Section 3 introduces our formal model of linguistic multilayer networks and the multiplex topic networks derived from them. Section 4 describes our experiments in detail, and Section 5 discusses our findings. Finally, Section 6 concludes and gives an outlook on future work.

2. Related Work

Our work is related to linguistic research on Tobler's [27] first law (TFL) which says that "[. . .] *everything is related to everything else, but near things are more related than distant things*" ([27], p. 236). Due to its underspecification, this so-called law raised many questions about what it means to be *related* or *distant* [28]. Accordingly, a range of approaches exist that make different proposals to interpret relatedness also in terms of *semantic relatedness*. In the context of information visualization, Montello et al. [29] test a variant of TFL called the first law of *cognitive* geography which says that "*people believe closer things to be more similar than distant things*" ([29], p. 317), where spatial distance is referred to for judging the similarity of information objects. This approach is contrasted with a study by Hecht and Moxley [30] who model relations of Wikipedia articles as a function of the probability of being linked in the web graph and find that this probability is related to the geographical distance of toponyms described in the articles. Hecht and Moxley relate their finding to the transitivity of networks by stating that the smaller the geographical distance of nodes, the higher their clustering coefficient ([30], 101). This work is extended by Li et al. [31], who calculate semantic relationships of articles instead of hyperlinks and show that TFL holds independently of the geographical domain up to a certain distance threshold. A lexical variant of TFL is mentioned by Yang et al. [32], according to which geographically close words tend to be clustered into the same geographical topics. This phenomenon has earlier been studied by Louwerse et al. (cf. the review in [26]) who reformulated Firth's famous dictum by saying that "[. . .] *you shall know the physical distance between locations by the lexical company they keep*" ([26], p. 1557). This means that the distance of places correlates with syntagmatic associations between the lexical items used to describe them. That is, language encodes geographical information [16] at least regarding the distances of semantically related places. From this perspective, TFL appears to be reformulated as a candidate for a geolinguistic law that is compatible with the more general *Symbol Interdependency Hypothesis* (SIH) [33]. According to SIH, linguistic information encodes perceptual

information so that the former serves as a shortcut to the latter [33]. Finally, a rather text-linguistic variant of TFL is proposed by Adams and McKenzie [34], which states that near places are each described by texts whose topics are more similar than in the case of texts about distant places.

In contrast to these approaches, we hypothesize that places, no matter how far apart, have similar topic distributions when their descriptions are transmitted by media such as city and region wikis. If we find evidence for this hypothesis, there are various candidates for explaining it: Firstly, such a finding could indicate a trivial meaning of TFL (cf. [28]) in relation to the topics modeled by us, implying that everything, distant or not, is highly related. Secondly, it could indicate the (in)effectiveness of distances and similarities at different scales: at the level of local, specific topics (within the scope of TFL) and at the level of global, more general topics (outside the scope of TFL). Thirdly, such a finding could indicate a hidden similarity of processes of collaboratively writing wikis about different places, even if the wikis are written by different communities (see Hypothesis 1). In order to decide between these alternatives, we need a new topic model that derives networks of thematic structures at different scales from texts in online media about the same places. This should at least include the networking of topics along relations of intertextuality and coauthorship in order to allow for revealing similarities of the underlying processes of collaborative writing. To this end, we will develop multiplex networks that integrate text- and author-driven topic networks.

So far, most approaches to thematic aspects of places use topic modeling based on *Latent Dirichlet Allocation* (LDA) to associate topics and texts about geographical units, where topics are represented as sets of thematically related words. An early approach in this regard is described by Mei et al. [35] who model *spatiotemporal theme patterns* to identify dominant topics in texts that are connected to places. A related approach is proposed by Qiang et al. [36], who aim to detect topics that are "localized" in places. This is done to ground their similarities in relations of their thematic representations—a scenario that is omnipresent in linguistically motivated work in the context of TFL (cf. Figure 1). Likewise, Adams and McKenzie [34] extract topic models from travel blogs to detect topics as groups of semantically related words associated to places, so that relations among places can be identified by shared topics. Another example is proposed by Bahrehdar and Purves [37]: instead of documents written by individual authors, they analyze tagging data extracted from image descriptions in Flickr. A hybrid model of topic modeling comes from Yin et al. [38], in which representations of regions are used instead of documents to link topics to places. A related *region-topic model* that uses regions as topics to map words, sentences, and texts to distributions of regions or to ground them semantically (cf. [39]) is proposed by Speriosu et al. [40]. A promising extension is developed by Gao et al. [41] who aim at detecting higher-level functional regions as semantically coherent areas of interest. To this end, they analyze co-occurrence relations between topics to describe many-to-many relations of locations and urban functions. Another direction is pursued by Lansley and Longley [42], who investigate the location- and time-based distribution of topics in Twitter, setting a number of twenty topics as a target for LDA. See also Jenkins et al. [13] who utilize a list of six high-level topic categories. One of the largest studies in this context is the one of Gao et al. [43] who present an integrative approach to modeling texts from a range of different media such as Wikipedia, Twitter, and Flickr to demarcate cognitive regions [44]. All these approaches start from topic modeling to map natural language texts onto distributions of topics in order to relate the places thematized by these texts (cf. Figure 1).

A prominent precursor of topic models [45] is given by *Latent Semantic Analysis* (LSA) [46]. Consequently, there are studies in the context of TFL based on this predecessor. Davies [24], for example, interprets the associations of place names computed by LSA from place descriptions as a model of the cognitive representation of the corresponding spaces (cf. [47]). This approach opens up a perspective for measuring biased cognitive representations of spatial systems: according to Davies, her approach provides representations of cognitive geographies that are explored by the associations of semantically close place names in accordance or not with the underlying geographical relations, that is, in accordance or not with TFL (cf. [39]). These and related studies produce interesting results about the localization of topics or vice versa about the thematization of places in texts. However, they mostly disregard topic networking, not to mention the networking of topics viewed from different angles. Although it is easy to derive a network approach from binary relations of topic similarity, relationships that cannot be traced back to sharing similar words are hardly mapped by topic models of the sort considered so far. By generating topic distributions per location, for example, we know nothing about the dynamics of the coauthorship of the underlying texts: in the extreme case, one observes (dis) similarities, which result from the activity of a small number of authors or even only one author—in contrast to the assumed collaboration density of online media such as Wikipedia. Therefore, it is our goal to develop a model of topic networks that simultaneously addresses the dynamics of the coauthorship of the underlying texts. A subtask will be to develop a formal model of thematic networking that is generic enough to integrate a wide range of sources of networking—at least theoretically.

While most of the approaches considered so far ignore aspects of networking, a second branch of research tends to follow the paradigm of network theory. Hu et al. [48], for example, measure the semantic relatedness of cities as nodes of a city network [9] depending on the co-occurrences of city names in news articles. This approach is related to Liu et al. [49], who explore co-occurrences of toponyms to induce city networks that can be used to test predictions associated with TFL. Hu et al. [48] further develop this approach to networking cities by reference to topics of articles in which the corresponding toponyms are observed. They use Labeled LDA [50] to learn to extract topics α from texts to finally determine the α-relative similarity of cities based on the co-occurrences of their names in texts about α. Another

approach to city networks using Wikipedia as a data source is proposed by Salvini and Fabrikant [9]: they link cities as a function of the number of articles "co-siting" [51] their Wikipedia articles. A comprehensive perspective on modeling spatial information is developed by Luo et al. [52], who propose a three-part network model that integrates representations of spatial, social, and semantic networks. In this conceptual model, semantics plays the role of interpreting behavior in spatial and social space and thus of bridging them. Although we share this hybridization of the network perspective on spatial information, we strive for a more concrete model that can be empirically tested.

Any such study has to face various aspects of the vagueness [44, 53] or informational uncertainty [5] of concepts of regions [44] and places [13] and especially of the names of such entities [43]. According to Winter and Freksa [54], this includes *semantic ambiguity*, *indeterminacy of spatial extent*, or *boundary vagueness* [43], preference-oriented *re-scaling of extent*, and the *dynamics of salience* affected by various dimensions of contrast. Beyond boundary vagueness, Gao et al. [43] speak of the shape and location vagueness by example of cognitive regions. Furthermore, Jenkins et al. [13] refer to the temporal dynamics of places as evolving concepts as a source of uncertainty. From a methodological point of view, this multifaceted uncertainty has two implications: in relation to the model, which should be flexible enough to map these facets, and in relation to the object itself, which could complicate its modeling by unsystematically distorting it.

In accordance with Hu's study [55], we assume that the thematic perspective complements the spatial and temporal perspective of the study of places. A rheme can be understood as the "content" of a geographical region that expands its dimensionality [44]. This content may be further specified in terms of affordances, functions, or shared conceptual representations associated by members of a community with the corresponding place so that different places can be related by being associated with similar content. This thematic perspective will be at the core of our article. To this end, we follow the approach of Jenkins et al. [13], according to which places are connected with meanings generated by collaborators of crowd-sourcing media such as Wikipedia: their collaboration creates what Jenkins et al. call *platial themes*, namely, themes that are characteristic for certain places. As shared meanings, these platial themes ultimately create a "*collective sense of place*," as it is perceived by the corresponding community. In this context, Jenkins et al. [13] propose to study *politics*, *business*, *education*, *recreation*, *sports*, and *entertainment* as six high-level topics of places. However, by reference to the *Dewey Decimal Classification* (DDC), we will instead deal with more than six hundred hierarchically organized topics, each of which is manifested by a range of Wikipedia articles. In any event, we have to consider that thematic aspects may distort the conceptualization and perception of spatial objects [43]. A central question then concerns the regularity or systematicity of this distortion in the sense of asking to what extent thematic representations of different places show similar aspects of being biased. This question will be at the core of this article.

3. Multiplex Topic Networks: A Novel Approach to Topic Modeling

In order to study relations of thematic preference in VGI as a manifestation of distributed cognition, we introduce *Topic Network*s (TNs) as an alternative to *Topic Model*s (TMs) [23, 58, 59]. TMs are based on the idea that texts manifest probabilistic distributions of topics which are represented as probability distributions over the lexical constituents of these texts, where these distributions may be affected by style, the underlying genre, or any other (syntactic, semantic, or pragmatic) criterion of text production [60–62]. Regardless of its success, this model is unsuitable for modeling TNs as manifestations of distributed cognitive maps because of the following problems:

(P1) Corpus specificity: the corpus specificity of TMs impairs comparability and transferability to ever new corpora, since the topic distributions are learned from the input corpora whose topics are to be modeled. This approach apparently cannot use a transferable topic model as a basis for representing the topics of a large number of different corpora.

(P2) Topic labeling: the corpus-specific derivation of topic labels from the input corpora makes it difficult to compare their topic distributions. As reviewed by Herzog et al. [63], external resources can be used for this task. However, there are hardly any such resources for all possible topic combinations—unless one wants to explore an overarching system such as Wikidata making such a project considerably more difficult due to its size. The labeling problem can be addressed using, for example, Labeled LDA [50], an approach that leads us into the area of supervised classification, which is also followed here.

(P3) Scalability: instead of dealing with corpora of equally large texts, online communication often leads to sparse, tiny texts that sometimes consist of a single sentence, a single phrase, or a single word. Regardless of the size of the text, we need a procedure that determines its topic distributions so that texts of different sizes can be compared using topic models of comparable size. Even if small texts are postprocessed (after topic modeling) in such a way that their topic distributions are derived from their lexical constituents, such an approach would nevertheless mean to exclude text snippets from the training process.

(P4) Rare topics: one reason to prefer training by means of corpora as large as Wikipedia is to allow for detecting topics even if they form a kind of thematic hapax legomenon in the corpora to be analyzed. If we try to identify rare topics directly from these corpora, we will probably not detect them, since by definition these corpora do not provide enough information to identify such topics. In any event, the rarity of evidence about a topic should not be an impediment to identifying its occurrences even at the level of single sentences.

(P5) Methodical closeness: instead of deriving all distributions of all dependent and independent variables

as part of the same topic model, one possibly wants to include different information sources that are computed by different methods based on diverse computational paradigms (e.g., ontological approaches to measuring sentence similarities, approaches to word embeddings based on neural networks, and topic models). In order to enable this, we look for a methodologically open topic model that allows such different resources to be easily integrated.

In a nutshell, we are looking for an approach that (i) allows thematic comparisons of previously unforeseen text corpora using an underlying reference corpus, (ii) offers a generic solution to the problem of topic labeling, (iii) is highly scalable and can therefore map even the smallest text snippets to topic distributions, (iv) simultaneously takes rare topics into account, and (vii) is methodologically open and expandable. Such a topic network model is now developed in two steps: in Section 3.1, we introduce the underlying formal apparatus. This is done by deriving multiplex topic networks from linguistic multilayer networks. Section 3.2 describes a method by which this model is instantiated as a prerequisite for its empirical testing.

3.1. From Linguistic Multilayer Networks to Multiplex Topic Networks. In this section, we introduce multiplex topic networks. This is a type of network that is based on the idea of deriving the networking of topics of textual units by evaluating evidence from different sources of information such as text vocabulary, higher-level text components, distributed authorship or readership, genre, register, or medium. Since these sources of evidence can be explored in different compositions, this can lead to different perspectives on the salience and networking of the topics addressed by the same texts. Topic networks are multiplex precisely in this respect: the different evidence-providing perspectives may lead to different topic networks that allow comparisons to be made through which differences in the linguistic, social, or otherwise contextual embedding of thematizations become visible. This concept of a multiplex topic network is now being generically formalized.

To introduce multiplex topic networks, we start with defining *linguistic multilayer networks* (Definition 1) whose layeredness allows for distinguishing several (non)linguistic information sources of topic networking. We refer to supervised topic classifiers trained by means of large reference corpora to tackle the challenges P1, P2, P3, and P4. Based thereon, we introduce so-called *text topic networks* (Definition 3), which evaluate intra- and intertextual relations for the purpose of topic networking. Then, we introduce *two-level topic networks* (Definition 4) and exemplify them by *author* (Definition 5) and *word topic networks* (Definition 6), which explore relations of (co)authorship and lexical relatedness, respectively, as sources of topic networking. These notions are generalized to arrive at *n-level topic networks* (Definition 7) which are based on $n > 1$ informational sources of topic networking (cf. challenge P5). Finally, *multiplex topic networks* are defined as families of n-level topic networks (Definition 8) representing the networking of the same set of topics from different informational perspectives and thus allowing for mapping the thematic dynamics, for example, of descriptions of the same place.

Definition 1. Let $X = \{x_1, \ldots, x_n\}$ be a corpus of texts and $l \in \mathcal{N}, l > 1$. A *Linguistic Multilayer Network* (LMN) is a tuple (Mehler [57] speaks of *multilevel graphs*; see Boccaletti et al. [64] for a comprehensive overview of related notions whose formalism is used here; see Stella et al. [65] for an example of a multiplex network of lexical systems)

$$\begin{aligned}\mathscr{L}(X, l) &= (\mathbb{L}, \mathbb{C}),\\ \mathbb{L} &= \{L_i = (V_i, A_i, \mu_i, \nu_i, \lambda_i, \kappa_i) \mid i = 1, \ldots, l\},\\ \mathbb{C} &= \{C_{i.j} = (V_{i.j}, A_{i.j}, \mu_{i.j}, \nu_{i.j}, \lambda_{i.j}, \kappa_{i.j}) \mid i, j = 1, \ldots, l : i \neq j\},\end{aligned} \tag{1}$$

of two sets of directed graphs such that the set of *kernel layers* $\mathbb{L}$ consists of a pivotal text layer and several derivative layers, that is, a coauthoring layer, a language-systematic word layer, and possibly several layers modeling the networking of constituents of the pivotal texts:

(1) The pivotal text layer $L_1 = (V_1, A_1, \mu_1, \nu_1, \lambda_1, \kappa_1)$, also called text network, is spanned by texts of the corpus $V_1 = X$ such that A_1 is manifesting intratextual (as in the case of reflexive arcs) or intertextual relations

(2) The author layer $L_2 = (V_2, A_2, \mu_2, \nu_2, \lambda_2, \kappa_2)$, also called agent network, is spanned by the network of agents (co)authoring the texts in V_1 and their social relations

(3) The lexicon layer $L_3 = (V_3, A_3, \mu_3, \nu_3, \lambda_3, \kappa_3)$, also called word network, is spanned by the language-systematic lexical signs (i.e., lexemes and related units) used by agents of V_2 as part of their agent lexica to author the texts in V_1

(4) For $3 < i \leq l' < l$, $L_i = (V_i, A_i, \mu_i, \nu_i, \lambda_i, \kappa_i)$ is called a constituent layer modeling the networking of (e.g., lexical, phrasal, and sentential) constituents of texts $x \in V_1$ such that A_i maps intratextual (e.g., anaphoric) or intertextual (e.g., sentence similarity) relations

(5) For $l' < i \leq l$, $L_i = (V_i, A_i, \mu_i, \nu_i, \lambda_i, \kappa_i)$ is called a contextual layer modeling the networking of units (e.g., media, genres, and registers [66]) of the contextual embedding of texts $x \in V_1$ such that A_i maps, for example, relations of the switching, merging, or embedding [67, 68] of these contextual units

(6) For each $i, j \in \{1, \ldots, l\}, i \neq j, C_{i.j} \in \mathbb{C}, |\mathbb{C}| = l(l-1)$, is called a margin layer where $V_{i.j} = V_i \cup V_j$, $A_{i.j} \subseteq V_i \times V_j$, $\mu_{i.j} = \mu_i \cup \mu_j$, and $\lambda_{i.j} = \lambda_i \cup \lambda_j$.

For $i, j = 1, \ldots, l, i \neq j$, μ_i and $\mu_{i.j}$ are vertex weighting functions, ν_i and $\nu_{i.j}$ are arc weighting functions, λ_i and $\lambda_{i.j}$ are vertex labeling functions, and κ_i and $\kappa_{i.j}$ arc labeling functions. We say that the linguistic multilayer network $\mathscr{L}(X, l)$ is *spanned over the text corpus* X and *layered into* l *layers*.

Example 1. To illustrate our definitions, we construct a minimized example. Suppose a corpus of four texts $V_1 = X = \{x_1, x_2, x_3, x_4\}$, each containing three lexemes

$x_1 = \{w_1, w_2, w_3\}$, $x_2 = \{w_1, w_2, w_4\}$, $x_3 = \{w_5, w_6, w_7\}$, and $x_4 = \{w_4, w_8, w_9\}$ (for reasons of simplicity, we exemplify texts as bag-of-words), that is, $V_3 = \{w_1, \ldots, w_9\}$, $V_{3.1} = \{w_1, \ldots, w_9, x_1, \ldots, x_4\}$, and $A_{3.1} = \{(w_1, x_1), (w_2, x_1), (w_3, x_1), \ldots, (w_4, x_4), (w_8, x_4), (w_9, x_4)\}$. Further, we assume four authors $V_2 = \{a_1, a_2, a_3, a_4\}$ such that a_1 and a_2 coauthored x_1 and x_2, while a_3 and a_4 coauthored x_3 and x_4; that is, $V_{2.1} = \{a_1, \ldots, a_4, x_1, \ldots, x_4\}$ and $A_{2.1} = \{(a_1, x_1), (a_2, x_1), (a_1, x_2), (a_2, x_2), (a_3, x_3), (a_4, x_3), (a_3, x_4), (a_4, x_4)\}$. Further, we assume that the texts x_1 and x_2 are linked by some intertextual coherence relations (e.g., by a rhetorical relation, an argument relation, or some hyperlinks) as are the texts x_3 and x_4 so that $A_1 = \{(x_1, x_2), (x_3, x_4)\}$. Note that additional arcs of the layers L_1, L_2, and L_3 will be generated according to the subsequent definitions. For simplicity reasons, we assume all weighting functions to be limited to the set $\{0, 1\}$ of vertex/arc weights. Since we assume no additional constituent layer, we get $l = 3$. Thus, any linguistic multilayer network $\mathscr{L}(X, 3)$ based on this setting is layered into three layers.

Throughout this paper, we use the following simplifying notation: for any graph $G = (V, A, \lambda)$ of order $|G| = |V|$, arc set $A \subseteq V^2$ of size $|A|$ and vertex labeling function λ, and any vertex $v \in V$, we write $\dot{v} = \lambda(v)$. Thus, for any two graphs G_i and G_j with vertex labeling functions λ_i and λ_j, for which $\lambda_i(v) = \lambda_j(w)$, $v \in V_i$ and $w \in V_j$, we can write $\dot{v} = \dot{w}$. Further, for any function $f : X \times Y \longrightarrow Z$, for which $f(x, y) = z$, we use the following alternative notations:

$$f(x, y) = z \Longleftrightarrow x \longrightarrow_f y = z \Longleftrightarrow x \xrightarrow{f} y = z \Longleftrightarrow f_y(x) = z. \tag{2}$$

Finally, for any function $f : Z^n \longrightarrow Z$, Z being any set, we introduce the following notation based on square brackets:

$$\begin{aligned} &f(\ldots, x \xrightarrow{f} y, y \xrightarrow{g} x \ldots) = z \\ &\Longleftrightarrow f\left[\ldots, x \underset{g}{\overset{f}{\longleftrightarrow}} y \ldots\right] = z \Longleftrightarrow f\left[\ldots, x_g \longleftrightarrow_f y \ldots\right] \\ &\qquad = z. \end{aligned} \tag{3}$$

To leave no room for ambiguity, we assume that expressions of the sort $x \longrightarrow_f y, y \longrightarrow_g x$ are replaced from left to right into expressions of the sort $x_g \longleftrightarrow_f y$. Henceforth, a structure such as $x \longrightarrow_f y$ will be called *information link*. Based on Definition 1, we start now with introducing text topic networks using the following auxiliary notion.

Definition 2. Let $\mathscr{C} = (V_{\mathscr{C}}, A_{\mathscr{C}})$ be a directed *Generalized Tree* (GT) according to Mehler [69, 70] representing a hierarchical topic structure, henceforth called *Reference Classification System* (RCS), that is spanned by kernel arcs which are possibly superimposed by upward, downward, lateral, sequential, external, or reflexive arcs. (See Figure 2 for an example of a GT. This notion is required since we may decide for using, for example, the category system of Wikipedia as an RCS, which spans a GT [70]). That is, vertices $t \in V_{\mathscr{C}}$ represent topics, while kernel arcs $(t, u) \in A_{\mathscr{C}}$ represent subordination relations according to which u is a thematic specialization of t. Let further, θ denote a *hierarchical text classifier* [71] taking values in $V_{\mathscr{C}}$ that has been trained, validated, and tested by means of a reference corpus $\mathscr{R}$. Let now $\mathscr{L}(X, l) = (\mathbb{L}, \mathbb{C})$ be a LMN spanned over the text corpus X and layered into l layers. We call the structure

$$\mathscr{S} = (\mathscr{C}, \theta, \mathscr{L}(X, l)), \tag{4}$$

a *Definitional Setting* for defining topic networks.

Example 2. Given the LMN of Example 1, the *Dewey Decimal Classification* (see Section 3.2), and the topic classifier θ of [72], which uses the DDC as its Reference Classification System $\mathscr{C}$, a definitional setting is exemplified by $(\mathrm{DDC}, \theta, \mathscr{L}(X, 3))$. More specifically, by t_1, t_2, t_3 we will denote three topic labels of the third level of the DDC so that $V_{\mathscr{C}} = \{\ldots, t_1, t_2, t_3, \ldots\}$. Note that by using the DDC as a reference classification, the generalized tree of Definition 2 is reduced to a tree (see Section 3.2 for more details).

Definition 3. Given a definitional setting $\mathscr{S} = (\mathscr{C}, \theta, \mathscr{L}(X, l))$ according to Definition 2, a *Text Topic Network* (TTN) is a vertex- and arc-weighted simple directed graph

$$T(L_1) = T(L_1, \{\}) = (V, A, \mu, \nu, \lambda, \kappa), \tag{5}$$

with vertex set V and arc set $A \subseteq V^2$ which is said to be *derived from* $\mathscr{S}$ and *inferred from* L_1 by means of the optional classifier $\theta^{\leftarrow}$ and the monotonically increasing functions $\alpha, \beta, \gamma, \delta : \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+$ if and only if $\forall v \in V$ and $\forall a = (v, w) \in A$:

$$\begin{aligned} \mu(v) &= \alpha\left(\sum_{x \in V_1} \beta\left(\theta(x, \lambda(v)), \theta^{\leftarrow}(\lambda(v), x)\right)\right) \\ &= \alpha\left(\sum_{x \in V_1} \beta\left(\theta_x(\dot{v}), \theta^{\leftarrow}_{\dot{v}}(x)\right)\right) \\ &= \alpha\left(\sum_{x \in V_1} \beta\left(x \xrightarrow{\theta} \dot{v}, \dot{v} \xrightarrow{\theta^{\leftarrow}} x\right)\right) \\ &= \alpha\left(\sum_{x \in V_1} \beta\left(x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}\right)\right) > 0, \end{aligned} \tag{6}$$

$$\begin{aligned} \nu(a) &= \gamma\left(\sum_{x, y \in V_1} \delta\left(\theta(x, \lambda(v)), \theta^{\leftarrow}(\lambda(v), x), \theta(y, \lambda(w)), \right.\right. \\ &\qquad \left.\left. \theta^{\leftarrow}(\lambda(w), y), \nu_1(x, y)\right)\right) \\ &= \gamma\left(\sum_{x, y \in V_1} \delta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, y \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{w}, x \xrightarrow{\nu_1} y\right]\right) > 0, \end{aligned} \tag{7}$$

where $\mu : V \longrightarrow \mathbb{R}^+$ is a vertex weighting function, $\nu : A \longrightarrow \mathbb{R}^+$ an arc weighting function, $\lambda : V \longrightarrow V_{\mathscr{C}}$ an injective vertex labeling function, $V_{\mathscr{C}}(V) = \{\lambda(v) \mid v \in V\} \subseteq V_{\mathscr{C}}$, and κ an injective arc labeling function. $T(L_1)$ is called a

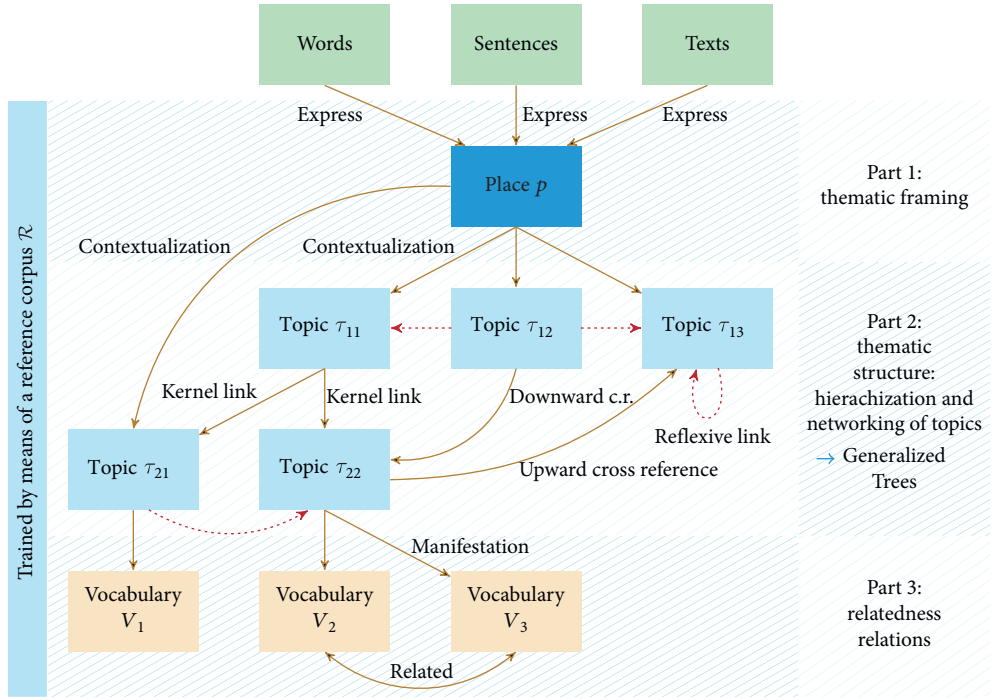


FIGURE 2: Schema of mapping texts onto hierarchically organized topic networks: words, sentences, and texts describing a certain thematic frame (e.g., a place as the central topic of a city wiki) are mapped onto a topic hierarchy as an example of a so-called generalized tree [56, 57]. Based on kernel links of thematic specialization, the topics are organized hierarchically, whereby this organization is superimposed by up- and downward cross-references. Dashed links are inferred as a result of modeling the thematic networking of input words, sentences, or texts. As we assume that the underlying topic model has been trained by means of a reference corpus $\mathcal{R}$ (see Definition 2), each topic is associated with a distribution of lexical elements of $\mathcal{R}$ that are preferably used to *manifest* this topic (see the types ν and w in relation to the topics α and β in Figure 1). This preference relation may be extended to higher-level units such as sentences.

one-layer topic network that is generated by the *generating layer* L_1.

Formulas (6) and (7) require that the weighting values for nodes and arcs are greater than 0: otherwise, the candidate vertices and arcs do not exist in the TTN. $\theta^{\leftarrow}$ is a classifier mapping pairs (t, x) of topics $t \in V_{\mathscr{C}}$ and texts x onto real numbers indicating the extent to which x is a "prototypical" instance of t (obviously, the textual arguments of the functions θ and $\theta^{\leftarrow}$ are not restricted to elements of X.)

Example 3. Given Example 2, we assume that $\lambda(\nu_1) = t_1$, $\lambda(\nu_2) = t_2, \lambda(\nu_3) = t_3$ and $\theta(x_1, t_1) = 1$, $\theta(x_2, t_2) = 1$, $\theta(x_3, t_3) = \theta(x_4, t_3) = 1$ so that $V = \{\nu_1, \nu_2, \nu_3\}$. In our example, we disregard $\theta^{\leftarrow}$. Further, we assume that the functions α, β, γ, and δ are identity functions. Thus, $\mu(\nu_1) = \mu(\nu_2) = 1$ and $\mu(\nu_3) = 2$. Now, we can generate a topic link between ν_1 and ν_2 by exploring the intertextual relation $(x_1, x_2) \in A_1$: To this end, we assume that

$$\delta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{\nu}, y \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{w}, x \overset{\nu_1}{\longrightarrow} y\right] \leftarrow \delta\left[x \overset{\nu_1}{\longrightarrow} y\right] \leftarrow \mathrm{id}\left(x \overset{\nu_1}{\longrightarrow} y\right) = x \overset{\nu_1}{\longrightarrow} y, \tag{8}$$

so that $\nu((\nu_1, \nu_2)) = 1$. By analogy to this case, we link topic ν_3 by means of a reflexive link so that $A = \{(\nu_1, \nu_2), (\nu_3, \nu_3)\}$.

Note that these simplifications are made for simplicity's sake only: Section 3.2 will elaborate a realistic weighting scenario. However, the function of the latter illustration is to show that by the intertextual linkage of both texts, we get evidence about the linkage of the topics instantiated by these texts. TTNs always operate according to this premise: they network topics as a function of the networking of an underlying set of texts. Figure 3 gives a schematic depiction of this scenario, which is varied subsequently to illustrate the other types of topic networks developed in this paper.

A concrete example of a TTN that is derived from the articles of the so-called Dresden wiki (see Section 4.1) is depicted in Figure 4. It shows the highest weighted topics addressed by these articles and their (undirected) links. The TTN has been computed by means of the procedural model of Section 3.2. Evidently, the topic *Transportation; ground transportation* is most prominent in this wiki followed by the topic *Central Europe; Germany*. Most topics belong to the areas *transportation* (red), *geography and history* (turquoise), and *architecture* (gray) (for the color code, see Appendix). More examples of TTNs can be found in Figures 5–7.

Arguments of the sort $x \longrightarrow_{\theta} \dot{\nu}$ can be used to quantify evidence about text x as an instance of topic $\dot{\nu}$: the more the evidence of this sort, the higher possibly the impact of x in formula (6) and the higher possibly the final weight of ν. The adverb *possibly* refers to what is licensed by the parameters γ and δ. Arguments of the sort $x \longrightarrow_{\nu_1} y$, where $x \neq y$, can be

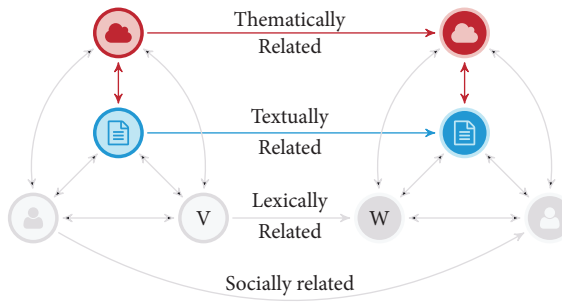


Figure 3: Schematic depiction of the informational sources of linking topics (red vertices) in text topic networks as a function of the textual relatedness of two texts (blue vertices) (belonging to layer L1 of a corresponding LMN—see Definition 1). Bidirectional red arcs denote arcs of the corresponding margin layers: in the present case, this concerns the relation between texts and topics (see below). Relations of thematic relatedness are inferred in this example (see Definition 3). Gray nodes and arcs indicate unconsidered sources of evidence.

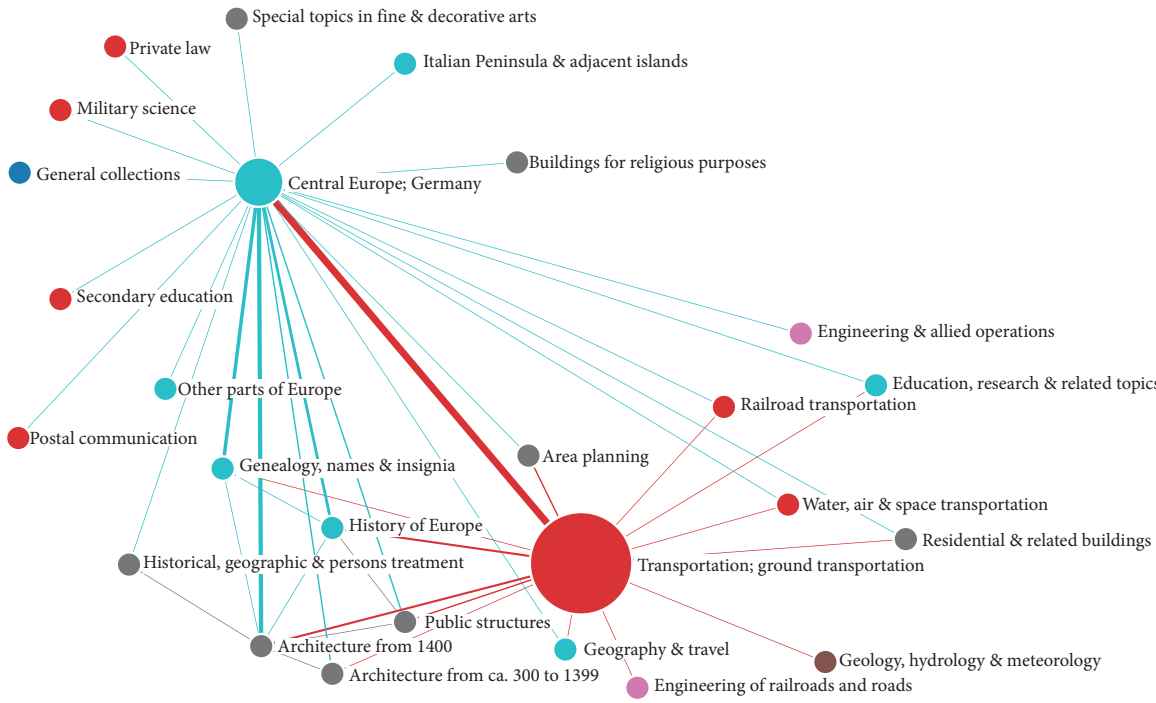


Figure 4: Visualization of a segment of the TTN of the city wiki Dresden (http://www.stadtwikidd.de/wiki/Hauptseite) using the 3rd level of the DDC as the underlying RCS for the definition of topics according to Section 3.2. The segment shows the highest weighted topics and their (undirected) links. Edges have been colored to show the two centers of this graph.

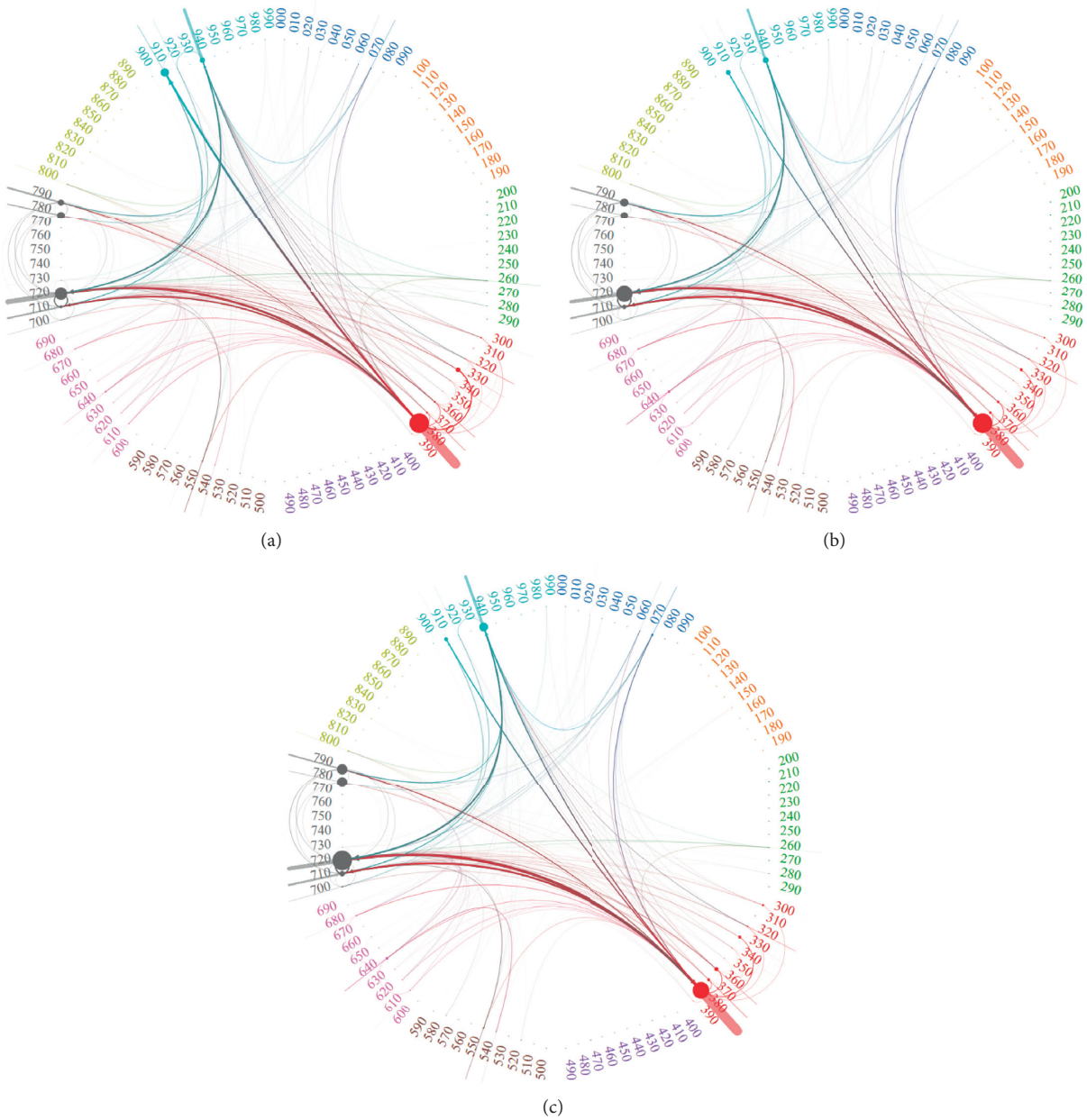
used to quantify evidence that text x is intertextually linked to text y: the more the evidence of this sort, the higher possibly the weight of the link from x to y and the higher possibly the influence of this link onto the weight of the link from topic v to topic w in formula (7) (in cases in which there is no explicit information about intertextual links, one can use functions of aggregated word embeddings of the lexical constituents of texts to calculate their intertextual similarity). In this and related definitions, we do not fully specify the functions $\theta, \theta^{\leftarrow}, \alpha, \beta, \gamma, \delta$ to leave enough space for different instances of topic networks.

Definition 3 relies on the pivotal text layer for deriving topic networks. To integrate further layers into the process of inferring topic networks, we introduce the following generalized schema.

Definition 4. Given a definitional setting $\mathcal{S} = (\mathscr{C}, \theta, \mathscr{L}(X, l))$ according to Definition 2, an $(L_1, \mathbb{L}')$-*Topic Network*, $\mathbb{L}' \in \{\varnothing\} \cup \{\{L_i\} \mid i \in \{2, \ldots, l\}\}$, is a vertex- and arc-weighted simple directed graph

$$T(L_1, \mathbb{L}') = (V, A, \mu, \nu, \lambda, \kappa), \tag{9}$$

which is said to be *derived from* $\mathcal{S}$ and *inferred from* L_1 and the elements of $\mathbb{L}'$ by means of the optional classifiers $\vartheta^{\leftarrow}, \vartheta : V_i \times V_{\mathscr{C}} \longrightarrow \mathbb{R}_0^+, \vartheta^{\leftarrow} : V_{\mathscr{C}} \times V_i \longrightarrow \mathbb{R}_0^+$ and monotonically increasing functions $\alpha, \beta, \gamma, \delta : \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+$ iff $\forall v \in V$ and $\forall a = (v, w) \in A$:

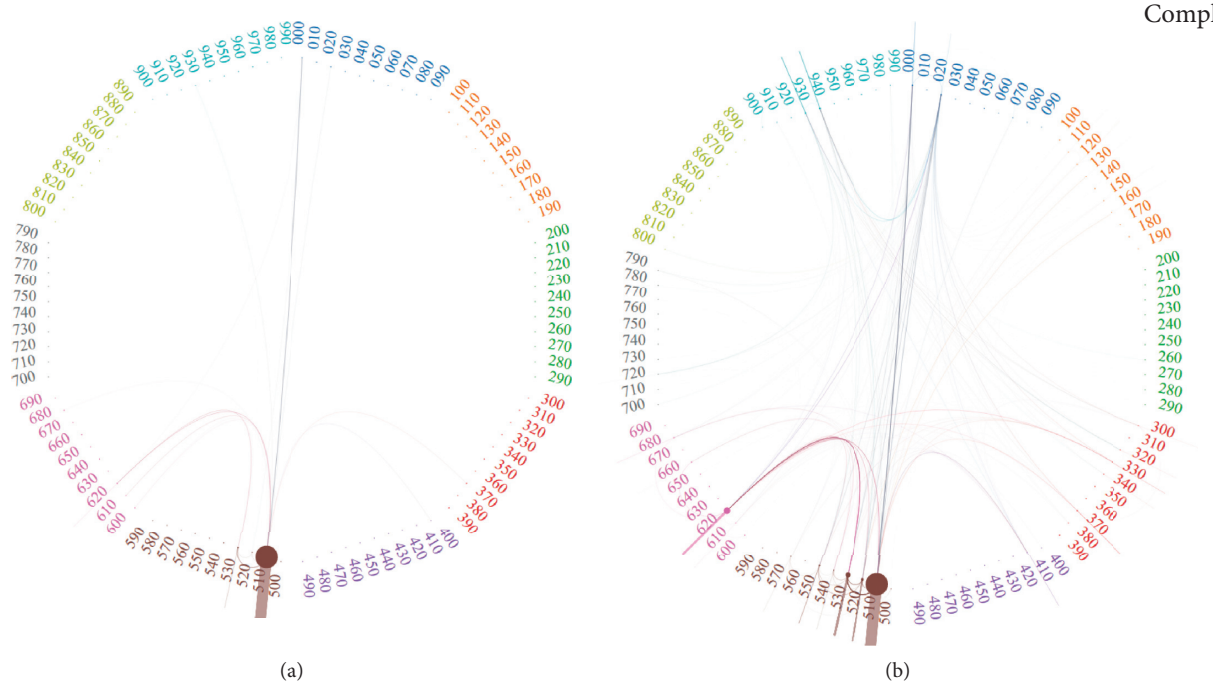


(a) (b) (c)

Figure 5: Visualizations of a TTN (top left) and two corresponding ATNs. The TNs are derived from the city wiki München (https://www.muenchenwiki.de/wiki/Hauptseite) (see Section 4 for statistics about this wiki) using the procedural model of Section 3.2. Top right shows the ATN for which (co)authorship activities are estimated by means of Wikipedia (see Section 3.2.3). The ATN for which these activities are estimated via the wiki itself is displayed below. The visualizations are carried out by means of PolyViz [73] regarding the 2nd level of the DDC: nodes are labeled (with numbers denoting the respective 2nd level class) and colored to encode their membership to one of the top 10 DDC classes (see Appendix). The higher the weight of a topic, the larger the node, and the higher the weight of an arc, the thicker the line. Node and line sizes are defined relative to the maximum vertex and arc weights of the underlying network.

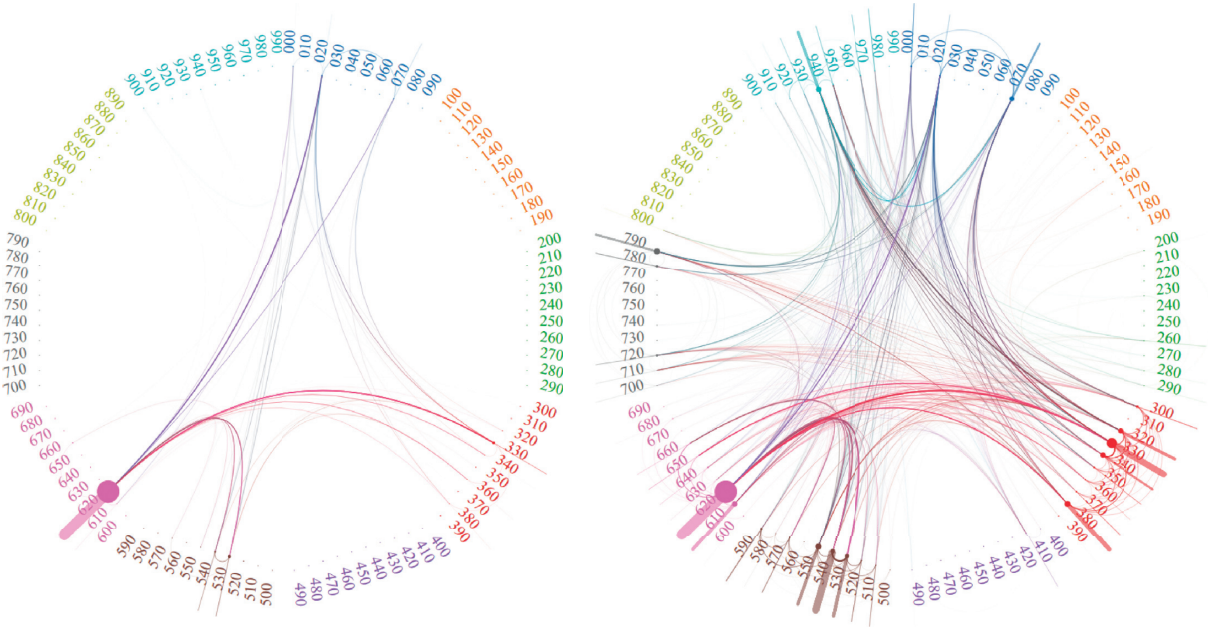
$$\mu(v) = \alpha\left(\sum_{x\in V_1, r\in V_i} \beta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, r \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{v}, r \underset{\nu_{1.i}}{\overset{\nu_{i.1}}{\longleftrightarrow}} x\right]\right) > 0. \tag{10}$$

$$\nu(a) = \gamma\left(\sum_{x,y\in V_1, r,s\in V_i} \delta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, y \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{w}, r \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{v}, s \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{w}, r \underset{\nu_{1.i}}{\overset{\nu_{i.1}}{\longleftrightarrow}} x, s \underset{\nu_{1.i}}{\overset{\nu_{i.1}}{\longleftrightarrow}} y, r \overset{\nu_i}{\longrightarrow} s, x \overset{\nu_1}{\longrightarrow} y\right]\right) > 0, \tag{11}$$



(a) (b)

Figure 6: Visualization (by means of PolyViz [73]) of the TTN of the 1st orbit (a) and of the 2nd orbit (b) of the German Wikipedia article *Integralrechnung* (*Integral*). The TTNs are derived from the corpora of articles in the 1st and the 2nd orbit (see formula (30)) of this article. Obviously, the most prominent 2nd level DDC class in both TTNs is 510 (*Mathematics*).



(a) (b)

Figure 7: Visualization (by means of PolyViz [73]) of the TTN of the 1st orbit (a) and of the 2nd orbit (b) of the German Wikipedia article *Kernkraftwerk* (*Nuclear power plant*). The TTNs are derived from the corpora of articles in the 1st and 2nd orbits (see formula (30)) of this article. Obviously, the most prominent 2nd level DDC class in both TTNs is 620 (*Engineering*). Compared to the example in Figure 6, the 2nd orbit is now thematically much more diversified.

where $\mathbb{L}' = \{L_i\}$. $\mu : V \longrightarrow \mathbb{R}^+$ is a vertex weighting function, $\nu : A \longrightarrow \mathbb{R}^+$ an arc weighting function, $\lambda : V \longrightarrow V_{\mathscr{C}}$ an injective vertex labeling function, $V_{\mathscr{C}}(V) = \{\lambda(v) \mid v \in V\} \subseteq V_{\mathscr{C}}$, and κ an injective arc labeling function. For $\mathbb{L}' = \{L_i\}$, we say that $T(L_1, \mathbb{L}')$ *is a two-level topic network* that is generated by the *generating layers* L_1 and L_i. If $\mathbb{L}' = \varnothing$, then formula (10) changes to formula (6) and formula (11) to formula (7). By omitting any optional classifier $g \in \{\theta^{\leftarrow}, \vartheta^{\leftarrow}\}$,

expressions of the sort $r_g \longleftrightarrow_f \dot{v}$ change to $r \longrightarrow_f \dot{v}$. ϑ is treated analogously.

To understand formula (10) look at Figure 8: among other things, formula (10) collects the triangle spanned by v, x, and a supposed that the two-level topic network is based on text and authorship links. Obviously, Definition 4 generalizes Definition 3. Now, it should be clear why we speak of the text network of an LMN as its pivotal level: it is the reference layer of any additional layer that is integrated into a two-level topic network according to Definition 4. This role is maintained below when we generalize this definition to capture n layers, $n > 2$. With the help of Definition 4, we can immediately derive so-called *author topic networks*.

Definition 5. An *Author Topic Network* (ATN) is a directed graph

$$T(L_1, \mathbb{L}') = (V, A, \mu, \nu, \lambda, \kappa), \tag{12}$$

according to Definition 4 such that $\mathbb{L}' = \{L_2\}$.

The relational arguments of this definition can be motivated as follows—assuming that they are instantiated appropriately:

(1) $x \xrightarrow{\theta} \dot{v}$ can be used to represent evidence that text x is about topic $\dot{v}$ possibly in relation to other topics of $V_{\mathscr{C}}$.

(2) $\dot{v} \xrightarrow{\theta^{\leftarrow}} x$ can be used to represent evidence that text x is a prototypical instance of topic $\dot{v}$ possibly in relation to other texts in V_1.

(3) $r \xrightarrow{\vartheta} \dot{v}$ can be used to represent the extent to which agent r tends to write about topic $\dot{v}$ possibly in relation to other topics of $V_{\mathscr{C}}$.

(4) $\dot{v} \xrightarrow{\vartheta^{\leftarrow}} r$ represents evidence that agent r is a prototypical author writing about topic $\dot{v}$ possibly in relation to other agents in V_2.

(5) For $x \neq y$, $x \xrightarrow{\nu_1} y$ can be calculated to represent evidence about text x to be intertextually linked to text y (e.g., in the sense of linking contributions of different authors). Otherwise, if $x = y$, $x \longrightarrow_{\nu_1} y$ can be used to quantify evidence about x being intratextually structured.

(6) $r \xrightarrow{\nu_{2.1}} x$ can be used to quantify evidence about the role of agent r as an author of text x possibly in relation to other texts authored by r. Typically, $\nu_{2.1}$ is a function of the number of edit actions performed by r on x [74].

(7) $x \xrightarrow{\nu_{1.2}} r$ can be used to quantify evidence about the role of agent r as a prototypical author of text x possibly in relation to other authors of x. In the simplest case, $\nu_{2.1}$ is symmetric making $\nu_{1.2}$ obsolete.

(8) $r \xrightarrow{\nu_2} s$ represents evidence that agent r is a coauthor of or interacting with s. For instantiating ν_2, the literature knows a wide range of alternatives [74, 75] (which mostly concern symmetric measures of coauthorship). Note that we do not require that $r \neq s$.

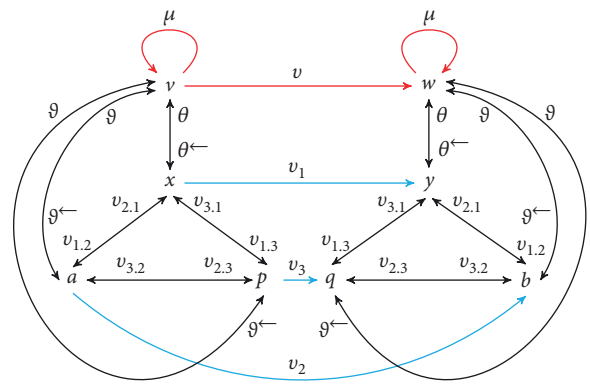


FIGURE 8: A diagrammatic depiction of inferred arcs (red) in topic networks, inferred by means of various arcs (black and blue) of an underlying LMN. Orientation of inferred arcs is provided by three types of input arcs (blue). $x, y \in V_1$ denotes two texts, $a, b \in V_2$ denotes two authors working on x, and y, respectively, $p, q \in V_3$ denotes two lexical units occurring in x and y, respectively. Inferred weights of vertices are denoted by means of (red) reflexive arcs.

Example 4. Starting from Example 3 to exemplify arcs between topics in *author topic networks*, we can now additionally explore the evidence, that text x_1 and x_2 are both coauthored by the agents a_1 and a_2. That is, we can assume a coauthorship link $(a_1, a_2) \in A_2$ (A_2 is the arc set of the author layer in Definition 1) of weight $\nu(a_1, a_2) = 1$. Let us now assume the following simplification of the function δ in Definition 4, for which we assume that it simply multiplies and adds up its argument values in the following way:

$$\begin{aligned}
&\delta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, y \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{w}, r \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{v}, s \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{w}, r \underset{\nu_{1.2}}{\overset{\nu_{2.1}}{\longleftrightarrow}} x, s \underset{\nu_{1.2}}{\overset{\nu_{2.1}}{\longleftrightarrow}} y, r \xrightarrow{\nu_2} s, x \xrightarrow{\nu_1} y\right] \longleftarrow \\
&\qquad \delta\left[x \xrightarrow{\theta} \dot{v}, y \xrightarrow{\theta} \dot{w}, r \xrightarrow{\nu_{2.1}} x, s \xrightarrow{\nu_{2.1}} y, r \xrightarrow{\nu_2} s, x \xrightarrow{\nu_1} y\right] \longleftarrow \\
&(x \xrightarrow{\theta} \dot{v}) \cdot (y \xrightarrow{\theta} \dot{w}) \cdot \left(r \xrightarrow{\nu_{2.1}} x\right) \cdot \left(s \xrightarrow{\nu_{2.1}} y\right) \cdot \left(r \xrightarrow{\nu_2} s + x \xrightarrow{\nu_1} y\right) = (1 \cdot 1 \cdot 1 \cdot 1)(1 + 1) \\
&\qquad = 2.
\end{aligned} \tag{13}$$

In our example, we get $\dot{v} = t_1 = \lambda(v_1)$, $\dot{w} = t_2 = \lambda(v_2)$, $x = x_1$, $y = x_2$, $r = a_1$, and $s = a_2$. Since there is no other interlinked pair of texts (see Example 1), instantiating the topics v_1 and v_2, we get $\nu((v_1, v_2)) = 2$ as the weight of this topic link in the corresponding ATN. By this simplified example of an ATN, we get the information that the link of topic v_1 to topic v_2 is additionally supported by the coauthorship of agents a_1 and a_2: this information extends the evidence about the topic link as provided by the underlying TTN of Example 3. Likewise, the reflexive link of topic v_3 is augmented by 1 compared to the underlying TTN, while there is no other topic link to be considered in this example of an ATN. By analogy to Figure 3, Figure 9 gives a schematic depiction of this scenario. Note that in our example, the weight of the link between authors a_1 and a_2 (cf. $r \xrightarrow{\nu_2} s$) is a function of their coauthorship: this is only one alternative to weight the social relatedness of both agents, actually one that can be measured by exploring (special) wikis. However, any other social relatedness might be explored to weight the interaction of agents.

By comparing a text topic network $T(L_1) = (V_{l+1}, A_{l+1}, \mu_{l+1}, \nu_{l+1}, \lambda_{l+1}, \kappa_{l+1})$ with an author topic network $T(L_1, \{L_2\}) = (V_{l+2}, A_{l+2}, \mu_{l+2}, \nu_{l+2}, \lambda_{l+2}, \kappa_{l+2})$ derived from the same LMN $\mathscr{L}(X, l)$, we can learn how the topics of $V_{\mathscr{C}}$ are manifested in the texts of corpus X in the form of a concomitance or a disparity of intertextual and coauthorship-based networking. Consider, for example, two vertices $v \in V_{l+1}$ and $w \in V_{l+2}$ such that $\dot{v} = \dot{w}$; let further $\perp$ and $\top$ denote the minimum and maximum that the vertex weighting functions of both graphs can assume. Then, we can distinguish four extremal cases:

(1) Cases of the sort

$$\perp \ll \mu_{l+1}(v) \approx \mu_{l+2}(w) \approx \top, \tag{14}$$

provide information on prominent topics that tend to be addressed by many texts which are coauthored by many authors.

(2) Situations like

$$\top \gg \mu_{l+1}(v) \approx \mu_{l+2}(w) \approx \perp, \tag{15}$$

probably apply to the majority of the topics in $V_{\mathscr{C}}$, which are hardly or even not at all addressed by texts in $V_1 = X$ due to the narrow thematic focus of these texts.

(3) Cases like

$$\top \approx \mu_{l+1}(v) \gg \mu_{l+2}(w) \approx \perp, \tag{16}$$

suggests a Zipfian topic effect, according to which a prominent topic is addressed by a small group of agents or even by a single author.

(4) Finally, situations of the sort

$$\perp \approx \mu_{l+1}(v) \ll \mu_{l+2}(w) \approx \top, \tag{17}$$

refer to rarely manifested topics addressed by a few but highly coauthored texts. In conjunction with many cases of the sort described by formula (16), situations of this kind indicate a Zipfian coauthoring effect, according to which many authors write only a few texts, while many texts are written by a few authors without encountering many (relevant) coauthors.

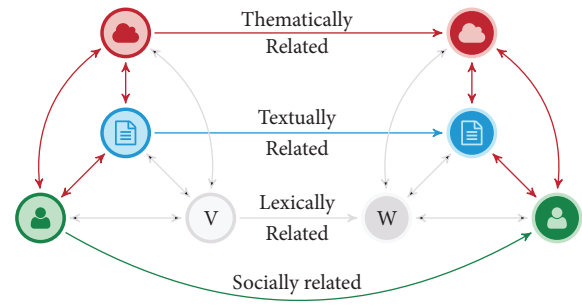


Figure 9: Schematic depiction of the informational sources of linking topics (red vertices) in *author topic networks* as a function of the textual relatedness of two texts (blue vertices) (that belong to layer L_1 of a corresponding LMN—see Definition 1) and the social relatedness of corresponding authors (green vertices) (that belong to layer L_2 of a corresponding LMN). Bidirectional red arcs denote arcs of the corresponding margin layers in Definition 1.

Formulas (14)–(17) compare the node weighting functions of a TTN with those of a related ATN. The same can be done regarding their arc weighting functions. That is, for two arcs $a = (r, s) \in A_{l+1}$ and $b = (v, w) \in A_{l+2}$, for which $\dot{r} = \dot{v} \wedge \dot{s} = \dot{w}$, we distinguish again four cases ($\perp$ and $\top$ now denote the minimum and maximum the arc weighting functions of both graphs can assume):

(1) In the case of

$$\perp \ll \nu_{l+1}(a) \approx \nu_{l+2}(b) \approx \top, \tag{18}$$

topic $\dot{v}$ is intertextually linked more strongly to topic $\dot{w}$ and authors of its text instances tend to cooperate with those of instances of topic $\dot{w}$ likewise to a greater extent.

(2) In the case of

$$\top \gg \nu_{l+1}(a) \approx \nu_{l+2}(b) \approx \perp, \tag{19}$$

topic $\dot{v}$ is intertextually less strongly linked to topic $\dot{w}$ and the few authors of its textual instances tend to cooperate with authors of instances of topic $\dot{w}$ likewise to a lesser extent.

(3) In the case of

$$\top \approx \nu_{l+1}(a) \gg \nu_{l+2}(b) \approx \perp, \tag{20}$$

topic $\dot{v}$ is intertextually more strongly connected with topic $\dot{w}$, while authors of its text instances tend to cooperate with those of instances of topic $\dot{w}$ to a lesser extent, if at all.

(4) Finally, in the case of

$$\perp \approx \nu_{l+1}(a) \ll \nu_{l+2}(b) \approx \top, \tag{21}$$

topic $\dot{v}$ is intertextually less strongly linked to topic $\dot{w}$, while the numerous authors of its text instances tend to cooperate with those of instances of topic $\dot{w}$ to a much greater extent.

Our central question regarding the relationship between TTNs and ATNs *derived from the same LMN* is whether these networks are similar or not. If they are similar, we expect that cases of the sort described by formulas (14), (15), (18), and (19) predominate so that cases matched by formula (14) are parallelized by those considered by formula (18) and where cases according to formula (15) are concurrent to those described by formula (19). An opposite situation would be that two topic nodes in the TTN are highly weighted but weakly linked, while they are weakly weighted but strongly linked in the corresponding ATN. In this case, a few or even only a single author is responsible for the thematic focus of the TTN. Note that this scenario reminds again of a Zipfian effect regarding the relation of TTNs and ATNs. By characterizing TTNs in relation to ATNs along these and related scenarios, we want to investigate laws of the interdependence of both types of networks, which may consist, for example, in the simultaneity of dense or sparse intertextuality-based networking on the one hand and dense or sparse coauthorship-based networking on the other. We may expect, for example, that the more related the two topics, the more likely the authors of their textual instances cooperate. However, not so much is known about such scenarios in the area of VGI especially with regard to Hypothesis 1. Thus, we address this gap at least by introducing a novel theoretical model which may help filling it.

Figure 5 exemplifies two ATNs in relation to a corresponding TTN (T1) which were computed using the apparatus of Section 3.2 to instantiate the formal model of this section. The upper right ATN (A1) is computed by globally weighting coauthorship activities based on Wikipedia (as explained in Section 3.2.3); the ATN (A2) below is calculated by weighting of these activities relative to the city wiki itself. Figure 5 shows that the topic with DDC number 720 (*Architecture*) is weighted higher in A1 than in T1. This is all the more pronounced in A2, where 720 becomes the most prominent topic and consequently displaces the top subject from T1, that is, topic 380 (*Commerce, communications* & *transportation*). That is, although topic 380 is most frequently addressed in this wiki's texts, topic 720 not only is almost as salient but also attracts many more activities among its interacting coauthors. Similar observations concern the switch of the roles of the topics 910 (*Geography* & *travel*) and 940 (*History of Europe*) from T1 to A1 and A2.

Regardless of the answer to this and related questions, we will also ask whether the shape of an ATN can be predicted if one knows the shape of the corresponding TTN and vice versa. To answer this question, we will consider LMNs of different text genres: of city wikis and regional wikis on the one hand and extracts of encyclopedic wikis on the other. We expect that LMNs spanned over corpora of the same genre exhibit a pattern of collaboration- and intertextuality-based networking that makes TTNs and ATNs derived from them mutually recognizable or predictable, whereas for LMNs generated from corpora of different genres this does not apply.

For reasons of formal variety, we now consider an alternative to author topic networks, namely, so-called word topic networks, which in turn are derived from Definition 4.

Definition 6. A *Word Topic Network* (WTN) is a directed graph

$$T\left(L_1, \mathbb{L}'\right) = (V, A, \mu, \nu, \lambda, \kappa), \tag{22}$$

according to Definition 4 such that $L' = \{L_3\}$.

This definition departs by five new relational arguments from Definition 5, which—if being instantiated appropriately—can be motivated as follows:

(1) $a \xrightarrow{\nu_{3.1}} x$ quantifies evidence about the role of word a as a lexical constituent of text x possibly in relation to all other texts in which a occurs. Typically, $\nu_{3.1}$ is implemented by a global term weighting function [76] or by a neural network-based feature selection function.

(2) $x \xrightarrow{\nu_{1.3}} a$ quantifies evidence about the role of the word a as a lexical constituent of the text x possibly in relation to other lexical constituents of x. Typically, $\nu_{1.3}$ is a local term weighting function, such as normalized term frequency [76], or a topic model-based function.

(3) $a \xrightarrow{\vartheta} \dot{v}$ represents evidence about the word a to be associated with the topic $\dot{v}$ possibly in relation to all other topics of $V_{\mathscr{C}}$.

(4) $\dot{v} \xleftarrow{\vartheta^{-}} a$ calculates evidence about the extent to which the topic $\dot{v}$ is prototypically labeled by the word a, possibly in relation to all other words in V_3.

(5) $a \xrightarrow{\nu_3} b$ quantifies evidence about the extent to which the word a associates the word b. Typically, ν_3 is computed by means of word embeddings [77].

Based on this list, we better understand what topic networks offer in contrast to TMs. This concerns the flexibility with which we can include informational resources computed by different methods (e.g., based on neural networks, topic models, and LSA) to generate topic networks (cf. challenge P5). Different relational arguments $X \longrightarrow_Z Y$ can be quantified using different methods, which in turn can belong to a wide range of computational paradigms. Table 2 gives an account of the generality of our approach by hinting at candidate procedures for computing the different relations of Figure 8.

Example 5. Starting from Example 3 to exemplify arcs between topics in *word topic networks*, we have to additionally explore evidence regarding the lexical relatedness of the vocabularies of the texts x_1 and x_2. In Example 1, we assumed that the intersection of both texts (represented as bags-of-words) is given by the set $\{w_1, w_2\}$. By analogy to

Example 4, we assume now the following simplification of the function δ of Definition 4:

$$\delta\Big[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, y \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{w}, r \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{v}, s \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{w}, r \underset{\nu_{1.2}}{\overset{\nu_{2.1}}{\longleftrightarrow}} x, s \underset{\nu_{1.2}}{\overset{\nu_{2.1}}{\longleftrightarrow}} y, r \xrightarrow{\nu_2} s, x \xrightarrow{\nu_1} y\Big] \longleftarrow (x \xrightarrow{\theta} \dot{v}) \cdot (y \xrightarrow{\theta} \dot{w}) \cdot \left(r \xrightarrow{\nu_{2.1}} x\right) \cdot \left(s \xrightarrow{\nu_{2.1}} y\right) \cdot \left(r \xrightarrow{\nu_2} s + x \xrightarrow{\nu_1} y\right). \tag{23}$$

In this scenario, we have to instantiate Definition 4 as follows: $\dot{v} \longleftarrow t_1 = \lambda(v_1)$, $\dot{w} \longleftarrow t_2 = \lambda(v_2)$, $x = x_1$, $y = x_2$, $r = w_1$, and $s = w_1$ for one summand and—everything else being constant—$r = w_2$ and $s = w_2$ for a second summand (for w_3 (w_4), we do not assume a lexical relatedness w.r.t. the words of text w_4 (w_3)). Note that under this regime, we assume that *relatedness of lexical constituents* only concerns shared usages of identical words—of course, this is a simplifying example. By analogy to the setting of Example 4, we have thus to conclude that $\nu((v_1, v_2)) = 4$ as the weight of the topic link from v_1 to v_2 in the corresponding WTN. For texts x_3 and x_4, we may alternatively assume that lexical relatedness does not only concern shared lexical items but also relatedness that is measured, for example, by means of a terminological ontology [83] or by means of word embeddings [77]. In this way, we may additionally arrive at a topic link between v_2 and v_3. In order to allow for a comparison of a WTN with its corresponding TTN, a more realistic weighting scheme is needed that also reflects above and below average lexical relatednesses of the lexical constituents of interlinked texts—in Section 3.2, we elaborate such a model regarding ATNs in relation to TTNs. Figure 10 gives a schematic depiction of the scenario of WTNs as elaborated so far.

It is worth emphasizing that instead of the (language-systematic) lexicon layer L_3, we may use a constituent layer $L_k, k > 3$, to infer a two-level topic network. For example, we can use the layer spanned by the sentences of the pivotal texts to obtain a sort of *sentence topic network*. In this case, $a \longrightarrow_{\nu_k} b$ may quantify evidence about the extent to which the sentence a entails the sentence b or the extent to which the sentence a is similar to the sentence b, etc., while $x \longrightarrow_{\nu_{1.k}} a$ may quantify evidence about the extent to which the sentence a is thematically central for the text x, etc. In sentence topic networks, topic linkage is a function of sentence linkage: prominent topics emerge from being addressed by many sentences, while prominent topic links arise from the relatedness of many underlying sentences. Another example of inferring two-level topic networks is to link topics as a function of places mentioned (by means of toponyms) within the texts of the underlying corpus X, where geospatial relations of these places can be explored to infer concurrent topic relations: if place p is mentioned in text x about topic $\dot{v}$ and place q in text y about topic $\dot{w}$, where the platial relation $R(p, q)$ relates p and q, this information can be used to link the topic nodes v and w in the corresponding topic network. As a result, we obtain networks manifesting the networking of topics as a function of parallelized geographical relations.

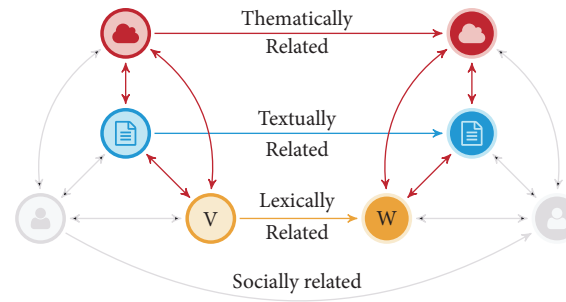


Figure 10: Schematic depiction of the informational sources of linking topics (red vertices) in *word topic networks* as a function of the textual relatedness of two texts (blue vertices) (that belong to layer L_1 of a corresponding LMN—see Definition 1) and the lexical relatedness of corresponding words (orange vertices) (that belong to layer L_3 of a corresponding LMN). Bidirectional red arcs denote arcs of the corresponding margin layers in Definition 1.

Obviously, any other relationship (e.g., entailment among sentences, sentiment polarities shared by linked texts, and co-reference relations) can be investigated to induce such two-level networks. And even more, we can think of n-level networks in which several such relationships are explored at once to generate topic links. We can ask, for example, which locations are linked by which geospatial relations while being addressed in which sentences about which topics where these sentences are related by which sentiment relations. Another example is to ask which authors prefer to write about which topics while tending to use which vocabulary: the higher the number of authors who use the same words more often to write about the same topic, and the higher the number of such words, the higher the weight of that topic. In this case, topic weighting is a function of frequently observed pairs of linguistic (here: lexical) means *and* authors. On the other hand, the higher the degree of coauthorship of two authors contributing to different topics and the higher the degree of association of the words used by these authors to write about these topics, the higher the weight of the link between the topics. This concept of a topic network induced by the text, the coauthorship, and the lexicon layer of an LMN is addressed by the following generalization, which provides a generation scheme for topic networks:

Definition 7. Given a definitional setting $\mathcal{S} = (\mathscr{C}, \theta, \mathscr{L}(X, l))$ according to Definition 2, an $(L_1, \mathbb{L}')$-*Topic Network*, for which

$$\mathbb{L}' = \{L_{i_1}, \ldots, L_{i_n}\} \in 2^{\{L_2, \ldots, L_l\}}, \tag{24}$$

is a vertex- and arc-weighted simple directed graph

$$T(L_1, \mathbb{L}') = (V, A, \mu, \nu, \lambda, \kappa), \tag{25}$$

which is said to be *derived from* $\mathcal{S}$ and *inferred from* L_1 and the elements of $\mathbb{L}'$ by means of the optional classifiers $\theta^{\leftarrow}, \forall i_j \in \{i_1, \ldots, i_n\} : \vartheta_{i_j} : V_{i_j} \times V_{\mathscr{C}} \longrightarrow \mathbb{R}_0^+, \vartheta_{i_j}^{\leftarrow} : V_{\mathscr{C}} \times V_{i_j} \longrightarrow \mathbb{R}_0^+$ and monotonically increasing functions $\alpha, \beta, \gamma, \delta : \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+$ iff $\forall v \in V$ and $\forall a = (v, w) \in A$:

$$\mu(v) = \alpha\left(\sum_{\substack{x\in V_1,\\ r_{i_1}\in V_{i_1},\ldots r_{i_n}\in V_{i_n}}} \beta\left[x \overset{\theta}{\underset{\theta^{\leftarrow}}{\longleftrightarrow}} \dot{v}, r_{i_1} \overset{\theta_{i_1}}{\underset{\theta^{\leftarrow}_{i_1}}{\longleftrightarrow}} \dot{v}, \ldots, r_{i_n} \overset{\theta_{i_n}}{\underset{\theta^{\leftarrow}_{i_n}}{\longleftrightarrow}} \dot{v}, r_{i_1} \overset{\nu_{i_1.1}}{\underset{\nu_{1.i_1}}{\longleftrightarrow}} x, \ldots, r_{i_n} \overset{\nu_{i_n.1}}{\underset{\nu_{1.i_n}}{\longleftrightarrow}} x\right]\right) > 0, \tag{26}$$

$$\begin{aligned}\nu(a) = \gamma\Bigg(\sum_{\substack{x,y\in V_1,\\ r_{i_1}\in V_{i_1},\ldots,r_{i_n}\in V_{i_n},\\ s_{i_1}\in V_{i_1},\ldots,s_{i_n}\in V_{i_n}}} \delta\Bigg[& x \overset{\theta}{\underset{\theta^{\leftarrow}}{\longleftrightarrow}} \dot{v},\ y \overset{\theta}{\underset{\theta^{\leftarrow}}{\longleftrightarrow}} \dot{w}, \\ & r_{i_1} \overset{\vartheta_{i_1}}{\underset{\vartheta^{\leftarrow}_{i_1}}{\longleftrightarrow}} \dot{v}, \ldots, r_{i_n} \overset{\vartheta_{i_n}}{\underset{\vartheta^{\leftarrow}_{i_n}}{\longleftrightarrow}} \dot{v}, s_{i_1} \overset{\vartheta_{i_1}}{\underset{\vartheta^{\leftarrow}_{i_1}}{\longleftrightarrow}} \dot{w}, \ldots, s_{i_n} \overset{\vartheta_{i_n}}{\underset{\vartheta^{\leftarrow}_{i_n}}{\longleftrightarrow}} \dot{w}, \\ & r_{i_1} \overset{\nu_{i_1.1}}{\underset{\nu_{1.i_1}}{\longleftrightarrow}} x, \ldots, r_{i_n} \overset{\nu_{i_n.1}}{\underset{\nu_{1.i_n}}{\longleftrightarrow}} x, s_{i_1} \overset{\nu_{i_1.1}}{\underset{\nu_{1.i_1}}{\longleftrightarrow}} y, \ldots, s_{i_n} \overset{\nu_{i_n.1}}{\underset{\nu_{1.i_n}}{\longleftrightarrow}} y, \\ & r_{i_1} \overset{\nu_{i_1}}{\longrightarrow} s_{i_1}, \ldots, r_{i_n} \overset{\nu_{i_n}}{\longrightarrow} s_{i_n}, \\ & r_{i_1} \overset{\nu_{i_1.i_2}}{\underset{\nu_{i_2.i_1}}{\longleftrightarrow}} s_{i_2}, \ldots, r_{i_1} \overset{\nu_{i_1.i_n}}{\underset{\nu_{i_n.i_1}}{\longleftrightarrow}} s_{i_n}, \ldots, r_{i_n} \overset{\nu_{i_n.i_1}}{\underset{\nu_{i_1.i_n}}{\longleftrightarrow}} s_{i_1}, \ldots, r_{i_n} \overset{\nu_{i_n.i_{n-1}}}{\underset{\nu_{i_{n-1}.i_n}}{\longleftrightarrow}} s_{i_{n-1}}, \\ & x \overset{\nu_1}{\longrightarrow} y\Bigg]\Bigg) > 0,\end{aligned} \tag{27}$$

$\mu : V \longrightarrow \mathbb{R}^+$ is a vertex weighting function, $\nu : A \longrightarrow \mathbb{R}^+$ an arc weighting function, $\lambda : V \longrightarrow V_{\mathscr{C}}$ an injective vertex labeling function, $V_{\mathscr{C}}(V) = \{\lambda(v) \mid v \in V\} \subseteq V_{\mathscr{C}}$, and κ an injective arc labeling function. For $|\mathbb{L}'| = n$, we say that $T(L_1, \mathbb{L}')$ *is an m-level,* $m = n + 1$, *topic network* generated by the *generating layers* L_1 and the elements of $\mathbb{L}'$. If $\mathbb{L}' = \varnothing$, formula (26) changes to formula (6) and formula (27) to formula (7). By omitting the optional classifier $g \in \left\{\vartheta^{\leftarrow}_{i_j} \mid j \in \{1, \ldots, n\}\right\}$, expressions of the sort $r_g \longleftrightarrow_f \dot{v}$ change to $r \longrightarrow_f \dot{v}$. θ and ϑ_{i_j} are treated analogously. In order to derive an *undirected* m-level topic network $\overline{T}(L_1, \mathbb{L}') = (V, E, \mu, \overline{\nu}, \lambda, \overline{\kappa})$ from $T(L_1, \mathbb{L}')$, we define $\{v, w\} \in E \longleftrightarrow (v, w) \in A \vee (w, v) \in A$ and

$$\overline{\nu}(\{v, w\}) = \begin{cases} \zeta_1(\nu((v, w)), \nu((w, v))), & (v, w) \in A \wedge (w, v) \in A, \\ \zeta_2(\nu((v, w))), & (v, w) \in A \wedge (w, v) \notin A, \end{cases} \tag{28}$$

and where ζ_1 and ζ_2 are monotonically increasing functions.

Evidently, Definition 7 is a generalization of Definition 3 by considering higher numbers of generating layers. A schematic depiction of the scenario addressed by this definition is shown in Figure 11 by example of a 3-level topic network that explores evidence about topic linking starting from the text, the author, and the lexicon layer of Definition 1. Likewise, Figure 12 depicts an n-level topic network, $n > 3$, in which additional resources are explored beyond the word, author, and text level. Figure 8 illustrates more formally the inference process underlying Definition 7, and in particular of the arguments used. It illustrates the inference of an arc that connects two topics by exploring the links of the text, author, and lexicon layers of an underlying LMN. In this example, the blue and black arcs are evaluated to determine the weights of red arcs connecting the focal topic nodes. Blue arcs are used to orientate inferred arcs. We will not develop this apparatus further, nor will we empirically examine $n + 1$-layer topic networks for $n > 2$. Rather, the apparatus developed so far serves to demonstrate the generality, flexibility, and extensibility of our formal model.

In the above, we explained that one of the reasons for introducing a flexible and extensible formalism of topic networks is to compare topic networks derived from different layers (e.g., from the text layer on the one hand and the author layer on the other). In order to systematize this approach, we finally introduce the concept of a *multiplex topic network*, which is derived from the same or from different linguistic multilayer networks:

Definition 8. Given a definitional setting $\mathcal{S} = (\mathscr{C}, \theta, \mathscr{L}(X, l))$ according to Definition 2, a *Multiplex Topic Network* (MTN) is a k-layer network

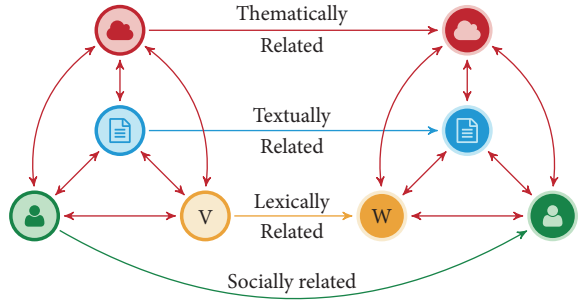


Figure 11: Schematic depiction of informational sources explored to link topics (red vertices) in a *3-level topic network* as a function of the textual relatedness of texts (blue vertices) (belonging to layer L_1 of Definition 1), the social relatedness of corresponding authors (green vertices) (belonging to layer L_2 of Definition 1), and the lexical relatedness of corresponding words (orange vertices) (belonging to layer L_3 of Definition 1). In this scenario, thematic relatedness is the information to be inferred, while textual, lexical, and social relations concern given information or evidence. Bidirectional red arcs denote arcs of corresponding margin layers of Definition 1.

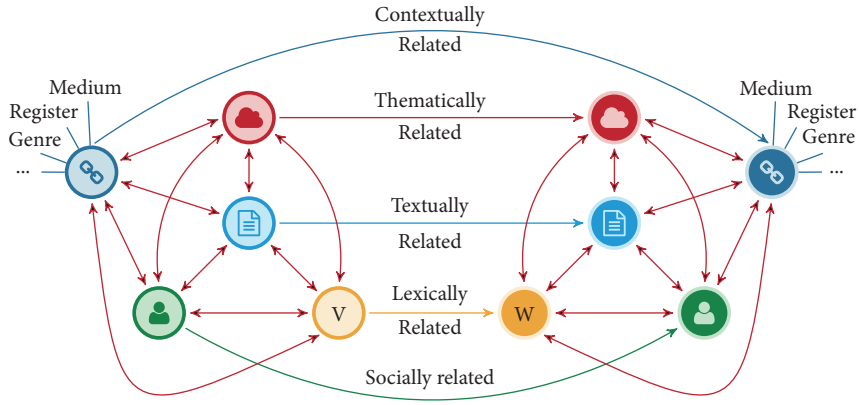


Figure 12: Schematic depiction of informational sources explored to link topics (red vertices) in an *n-level topic network*, $n > 3$, as a function of the textual relatedness of texts (blue vertices) (belonging to layer L_1 of Definition 1), the social relatedness of corresponding authors (green vertices) (belonging to layer L_2 of Definition 1), the lexical relatedness of corresponding words (orange vertices) (belonging to layer L_3 of Definition 1), and additional layers of contextual patterns concerning, for example, the underlying medium, genre, or register instantiated by the texts under consideration.

$$\begin{aligned}\mathscr{M}(X,k) &= (\mathbb{M},\mathbb{D}),\\ \mathbb{M} &= \{M_i = (V_i, A_i, \mu_i, \nu_i, \lambda_i, \kappa_i) \mid i = 1,\ldots,k\},\\ \mathbb{D} &= \{D_{i.j} = (V_{i.j}, A_{i.j}, \mu_{i.j}, \nu_{i.j}, \lambda_{i.j}, \kappa_{i.j}) \mid i,j = 1,\ldots,k : i \neq j\},\end{aligned} \tag{29}$$

such that each M_i, $i \in \{1,\ldots,k\}$, is an $(L_1, \mathbb{L}_i')$-Topic Network derived from $\mathcal{S}$ according to Definition 7 and for each $i,j \in \{1,\ldots,l\}$, $i \neq j$, $D_{i.j} \in \mathbb{D}$, $|\mathbb{D}| = k(k-1)$, is called a *margin layer* fulfilling the following requirements: $V_{i.j} = V_i \cup V_j$, $A_{i.j} = \{(v,w) \in V_i \times V_j \mid \dot{v} = \dot{w}\}$, $\mu_{i.j} = \mu_i \cup \mu_j$, and $\lambda_{i.j} = \lambda_i \cup \lambda_j$.

See Figure 13 for a schematic depiction of the comparison of two MTNs. Note that because of Definition 7, it does not necessarily hold that $V_{\mathscr{C}}(V_i) = V_{\mathscr{C}}(V_j)$, but it always holds that $V_{\mathscr{C}}(V_i) \subseteq V_{\mathscr{C}} \supseteq V_{\mathscr{C}}(V_j)$. In this respect, we depart from [64], which instead require more strongly that $V_i = V_j$. In the case of topic networks, this would be too restrictive, as different topic networks derived from the same definitional setting can focus on different subsets of topics, while ignoring the rest of the topics in the co-domain $V_{\mathscr{C}}$ of θ. (A way to extend Definition 8 is to include the RCS $\mathscr{C} = (V_{\mathscr{C}}, A_{\mathscr{C}})$ of Definition 2 as an additional layer. This would allow for directly relating its constituent topic networks with the hierarchical classification system $\mathscr{C}$.)

In this paper, we quantify similarities of the different layers of MTNs to shed light on Hypothesis 1. More specifically, we generate an LMN for each corpus of a set of different text corpora in order to derive a separate two-layer MTN for each of these LMNs, each consisting of a TTN and an associated ATN. Then, among other things, we conduct a triadic classification experiment: firstly with respect to the subset of all TTNs derived from our corpus, secondly with respect to the subset of all corresponding ATNs, and thirdly with respect to the subset of all TTNs in relation to the subset of the corresponding ATNs. In the next section, we explain the measurement procedure for carrying out this triadic classification experiment.

3.2. A Procedural Model of Topic Network Analysis. In order to instantiate topic networks as manifestations of the rhematic networking of places, we employ the procedure depicted in Figure 14. It combines nine modules for the induction, comparison, and classification of topic networks.

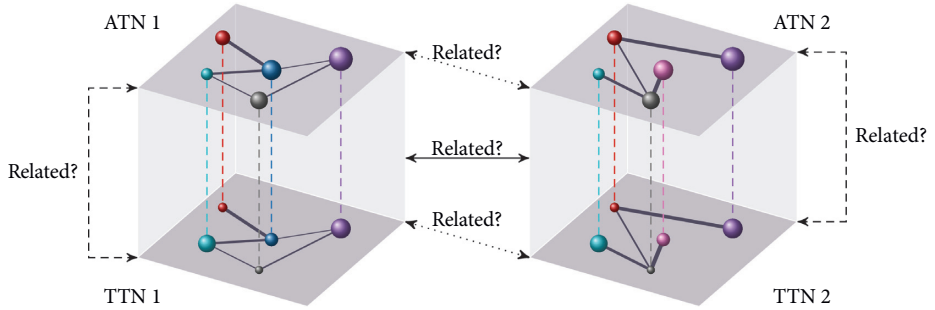


Figure 13: 3D depiction of two MTNs (left and right) each consisting of two layers (including a TTN at the bottom and an ATN at the top of the respective cube). Shared colors of nodes and dashed vertical lines indicate identically labeled vertices. The depiction disregards the orientation of the arcs. In this example, all four layers span topic networks over the same set of topics (vertices). Any such two-layer MTN can be used to represent the intertextuality- and coauthorship-based networking of the topics derived from the same corpus of texts about the same place. In this way, we gain several perspectives for the analysis of such multiplex networks: by comparing the TTNs or the ATNs of different MTNs (dotted arcs), by comparing the TTNs of different networks with their corresponding ATNs (dashed arcs), or by comparing the different MTNs as a whole with each other (solid arc).

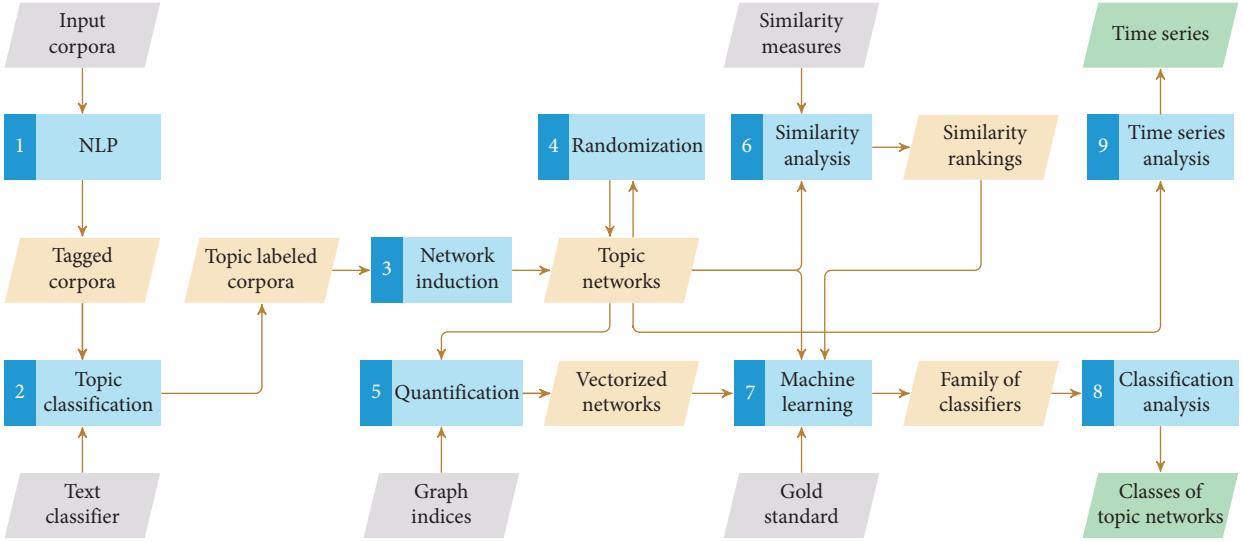


Figure 14: A procedural model of investigating LMNs and MTNs: generating, randomizing, and quantifying topic networks in 9 steps including *Natural Language Processing* (NLP) (1), topic classification using a classifier θ according to Definition 2 (2), topic network induction according to Definition 8 (3), network randomization according to Section 3.2.4 (4), network quantification (5), and network similarity analysis (6) both based on Section 3.2.6, machine learning of network classifiers (7) and classification analysis (8) both based on Section 3.2.7, and finally, time series analysis of topic networks (which will not be performed here) (9).

3.2.1. Module 1: Natural Language Processing. Preparatory for all modules is the natural language processing of the input text corpora. To this end, we utilize the NLP tool chain of *TextImager* [84] to carry out tokenization, sentence splitting, part of speech tagging, lemmatization, morphological tagging, named entity recognition, dependency parsing [85], and automatic disambiguation—the latter by means of fastSense [86]. For more details on these submodules, see [86, 87]. As a result of Module 1, the topic classification can be fed with texts whose lexical components are disambiguated at the sense level. As a sense model, we use the disambiguation pages of Wikipedia, currently the largest available model of lexical ambiguity.

3.2.2. Module 2: Topic Classification. According to Definition 2, the derivation of TNs from LMNs requires the specification of a *Reference Classification System* (RCS) $\mathscr{C} = (V_{\mathscr{C}}, A_{\mathscr{C}})$. For this purpose, we utilize the *Dewey Decimal Classification* (DDC), a system that is well established in the area of (digital) libraries. As a result, the generalized tree $\mathscr{C}$ from Definition 2 degenerates into an ordinary tree since the DDC has no arcs superimposing its kernel hierarchy (see Figure 15 for a subtree of the DDC). As a classifier θ, which addresses the DDC, we use $\theta :=$ text2ddc [72], a topic classifier based on neural networks, which has been trained for a variety of languages [88] (see https://textimager.hucompute.org/DDC/).

Starting from the output of Module 1 (NLP), we use text2ddc to map each input text x to the distribution of the 5 top-ranked DDC classes that best match the content of x as predicted by text2ddc. Since text2ddc reflects the three-level topic hierarchy of the DDC, this classifier can output a subset of 98 classes of the 2nd (two classes of this level are unspecified) and a subset of 641 classes of the 3rd DDC level for each input text. (We did not have training for all 3rd level classes (which are partly unspecified). See [72] and the appendix for details.) Thus, each topic network of each input corpus is represented on two levels of increasing thematic resolution. Note that text2ddc classifies input texts of any size (from single words to entire texts in order to meet challenge P3) and works as a multilabel classifier for processing thematically ambiguous input texts. By using an RCS, text2ddc meets challenge P2 simply by referring to the labels of the topic classes of the DDC. Furthermore, since text2ddc is trained with the help of a reference corpus, it can detect topics, even if they occur only once in a text (this is needed to meet challenge P4) and guarantees comparability for different input corpora (challenge P1). text2ddc is based on fastText whose time complexity is $O(h\log_2(k))$, where "k is the number of classes and h the dimension of the text representation" (2, [89]) (making this classifier competitive compared to TMs).

Figures 4–7 show examples of TTNs and ATNs generated by means of text2ddc by addressing the second level of the DDC. Each of these topic networks was generated for a subset of articles of the German Wikipedia that are at most 2 clicks away from the respective start article x (for the statistics of the corpora underlying these topic networks, see Section 4.1). Formally speaking, let $G = (V, A)$ be a directed graph and $v \in V$; the nth orbit induced by v is the subgraph,

$$\begin{aligned} G_v^n &= \left(V_v^n, A_v^n\right), \\ V_v^n &= \{w \in V \mid \delta(v,w) \leq n\}, \\ A_v^n &= \{(r,s) \in A \mid r, s \in V_v^n\}, \end{aligned} \tag{30}$$

that is induced by the subset of vertices whose geodetic distance $\delta(v,w)$ from v is at most n (cf. [90]). We compute the first orbit and the second orbit of a set of Wikipedia articles (so that G denotes Wikipedia's web graph). This is done to obtain a basis for comparison for the evaluation of topic networks derived from special wikis. Since Wikipedia is probably more strongly regulated than these special wikis, we expect higher disparities between networks of different groups (Wikipedia vs. special wiki) and smaller differences for networks of the same group.

3.2.3. Module 3: Network Induction. Network induction is done according to the formal model of Section 3.1. It starts with inducing an LMN $\mathscr{L}(X,2)$ for each input corpus X. That is, for each corpus X, we generate a text network L_1 and an agent network L_2 according to Definition 1:

(1) In this paper, X always denotes the set of texts (web documents) of a corresponding wiki W so that the text layer $L_1 = (V_1, A_1, \mu_1, \nu_1, \lambda_1, \kappa_1)$ of the LMN $\mathscr{L}(X,2)$, in which L_2 is an agent network defined below, can be used to represent the web graph [91] of this wiki. Thus, for any two texts x and y that are linked in W, we generate an arc $a = (v,w) \in A_1$, where $\nu_1(a) = 1$ and $\kappa_1(a) =$ hyperlink. Further, for $\forall x \in V_1 : \mu_1(x) = 1 \wedge \lambda_1(x) = x$.

(2) The author layer $L_2 = (V_2, A_2, \mu_2, \nu_2, \lambda_2, \kappa_2)$ of the LMN $\mathscr{L}(X,2)$ corresponding to L_1 (see Definition 1) is generated as follows: V_2 is the set of all registered authors or TCP/IP addresses of anonymous users working on texts in X so that $\forall v \in V_2 : \lambda_2(v)$ maps to this name or IP address, respectively. Let $✎(r,x)$ be the sum of all additions made by the author $r \in V_2$ to any revision of the edit history of the text x; we use $✎(r,x)$ to approximate the more difficult to measure concept of authorship as introduced by Brandes et al. [74]. Then, we define: $\forall r \in V_2 : \mu_2(r) = \sum_{x \in V_1} ✎(r,x)$. Further, A_2 is the set of all arcs (r,s) between users $r, s \in V_2$, for which there is at least one text x to which both contribute so that $✎(r,x)$, $✎(s,x) > 0$. Then, we define (cf. [92]):

$$\nu_2(r,s) = \sum_{x \in V_1} 2\frac{\min(✎(r,x), ✎(s,x))}{\sum_{u \in V_2} ✎(u,x)} \in (0,1]. \tag{31}$$

Finally, $\kappa_2(a) =$ coauthorship. Obviously, L_2 is symmetric.

Now, given the definitional setting $(\mathscr{C}, \theta, \mathscr{L}(X,2))$, where $\mathscr{C}, \theta$ are instantiated in terms of Section 3.2.2, we induce a TTN $T(L_1) = (V_{L_1}, A_{L_1}, \mu_{L_1}, \nu_{L_1}, \lambda_{L_1}, \kappa_{L_1})$ according to Definition 3 by means of appropriately defined monotonically increasing functions α_1, β_1, γ_1, and δ_1. To this end, we utilize the set

$$\theta_x^{V_{\mathscr{C}}} = \{\theta_x(\dot{v}) > \theta_{\min} \mid \dot{v} \in V_{\mathscr{C}}\}, \tag{32}$$

of the membership values of text $x \in V_1$ to the topics in $V_{\mathscr{C}}$, where the parameter $\theta_{\min}$ denotes a lower bound of an acceptable degree of aboutness. We set $\theta_{\min} := 0$. Further, by

$$\overline{\theta} = \frac{1}{|\mathbb{Y}|} \sum_{y \in \mathbb{Y}} y, \tag{33}$$

we denote the mean value of the set $\mathbb{Y} = \cup_{x \in V_1} \theta_x^{V_{\mathscr{C}}}$ of selected topic membership values and by $\max(\mathbb{X}, m)$ we denote the $m \in \{1, \ldots, |\mathbb{X}|\}$ largest value of the arbitrary set $\mathbb{X}$. Finally, we select a number $0 < m_\perp < |V_{\mathscr{C}}|$ and define $\forall v \in V, \forall x \in V_1$, thereby instantiating the parameters α, β, γ, and δ of formulas (6) and (7) of Definition 3:

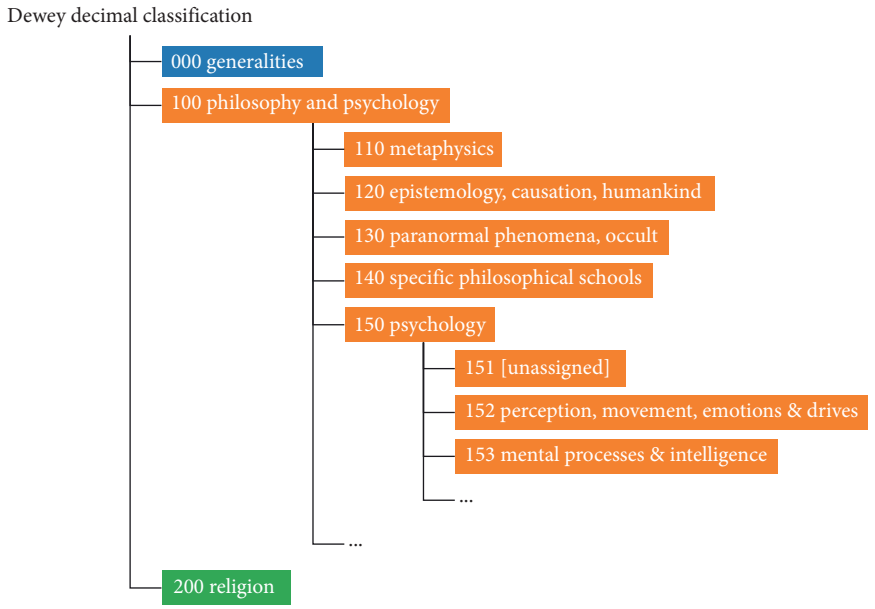


Figure 15: A subtree of the DDC displaying a snapshot of the second class (100) on the first three levels.

$$\alpha \coloneqq \alpha_1 = \mathrm{id}, \tag{34}$$

$$\begin{aligned}\beta\left(x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}\right) &\coloneqq \beta_1\left(x \xrightarrow{\theta} \dot{v}\right) = \beta_1\left(\theta\left(x, \lambda(v)\right)\right) = \beta_1\left(\theta_x(\dot{v})\right) \\ &= \begin{cases} \theta_x(\dot{v}), & \theta_x(\dot{v}) \in \left\{r \in \theta_x^{V_{\mathscr{C}}} \mid \exists m \leq m_{\perp} : r = \max\left(\theta_x^{V_{\mathscr{C}}}, m\right) \geq \overline{\theta}\right\}, \\ 0, & \text{else}, \end{cases}\end{aligned} \tag{35}$$

$$\begin{aligned}\gamma &\coloneqq \gamma_1 = \mathrm{id}, \\ \delta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, y \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{w}, x \xrightarrow{\nu_1} y\right] &\coloneqq \delta_1\left(x \xrightarrow{\theta} \dot{v}, y \xrightarrow{\theta} \dot{w}, x \xrightarrow{\nu_1} y\right) \\ &= \begin{cases} \beta_1\left(\theta_x(\dot{v})\right)\beta_1\left(\theta_y(\dot{w})\right), & (x, y) \in A_1, \\ 0, & \text{else}. \end{cases}\end{aligned} \tag{36}$$

According to formula (35), $\beta_1\left(x_{\theta^{\leftarrow}} \longleftrightarrow_{\theta} \dot{v}\right) = \theta_x(\dot{v})$ iff $\theta_x(\dot{v})$ is one of the $m_{\perp}$ highest membership values of x to the topics in $V_{\mathscr{C}}$, supposed that $\theta_x(\dot{v}) > \overline{\theta}$. Otherwise, $\beta_1\left(x_{\theta^{\leftarrow}} \longleftrightarrow_{\theta} \dot{v}\right) = 0$. In this paper, we experiment with $m_{\perp} = 5$. The higher the value of $m_{\perp}$, the more sensitive the generation of $T(L_1)$ to the thematic ambiguity of the underlying texts. However, since θ creates a membership value for each pair of texts and topics, we use $\overline{\theta}$ as a lower bound of aboutness (in the sense of addressing a topic known by θ) so that irrelevant classifications $\theta_x(\dot{v})$ do not affect $\mu_{L_1}(v)$.

Regarding the ATN $T(L_1, \{L_2\}) = (V_{L_2}, A_{L_2}, \mu_{L_2}, \nu_{L_2}, \lambda_{L_2}, \kappa_{L_2})$ corresponding to the TTN $T(L_1)$, we have to define monotonically increasing functions α_2, β_2, γ_2, and δ_2. To this end, we use several auxiliary functions:

(i) By $\overline{\not\in(\cdot,\cdot)}$, we denote the mean activity per author per Wikipedia article.

(ii) By $\overline{|\not\in(\cdot,\cdot)|}$, we denote the average number of active authors per Wikipedia article.

The corresponding estimators are found in Table 4. Now, consider the set $V_2(x)$ of all active authors of the text x and the set $\theta_v(V_1)$ of all texts that potentially contribute to $\mu_{L_2}(v)$ and thus to the weight of the vertex $v \in V_{L_2}$:

$$\begin{aligned}V_2(x) &= \{r \in V_2 \mid \not\in(r, x) > 0\}, \\ \theta_v(V_1) &= \left\{x \in V_1 \mid \beta_1\left(\theta_x(\dot{v})\right) > 0\right\}.\end{aligned} \tag{37}$$

Then, we define the following functions and ratios:

$$\mathrm{scale} = \begin{cases} (0,1]^2 \longrightarrow (0,2], \\ \mathrm{scale}(a,b) \mapsto 2\dfrac{a}{a+b}, \end{cases} \tag{38}$$

$$\omega_x = \mathrm{scale}\left(\left|V_2(x)\right|, \overline{|\not\in(\cdot,\cdot)|}\right), \quad \in (0,2], \tag{39}$$

$$\omega_v = \frac{1}{\left|\theta_v(V_1)\right|} \sum_{x \in \theta_v(V_1)} \omega_x, \quad \in (0,2], \tag{40}$$

where scale is a function which is used to rescale below or above average values (see formula (39)). Formula (40) defines the mean of the rescaled numbers of active users per article in $\theta_v(V_1)$. Based on these preliminaries and regarding the vertex weighting function μ_{L_2}, we define $\forall v \in V$ and $\forall r \in V_2$, thereby instantiating the functions α and β of formula (10) of Definition 4:

$$\alpha \coloneqq \alpha_2 \wedge \forall z \in \mathbb{R} : \alpha_2(z) = \omega_v \cdot z, \tag{41}$$

$$\beta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, r \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{v}, r \underset{\nu_{1,i}}{\overset{\nu_{i,1}}{\longleftrightarrow}} x\right] \coloneqq \beta_2\left(x \overset{\theta}{\longrightarrow} \dot{v}, r \overset{\nu_{2,1}}{\longrightarrow} x\right)$$
$$= \beta_1\left(\theta_x(\dot{v})\right) \cdot \begin{cases} \frac{1}{p} \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)}, & ✎(r,x) < \overline{✎(\cdot,\cdot)}, \\ \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)}, & ✎(r,x) = \overline{✎(\cdot,\cdot)}, \\ p \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)}, & ✎(r,x) > \overline{✎(\cdot,\cdot)}. \end{cases} \tag{42}$$

In the present paper, we experiment with $p = 2$. To understand this definition, we have to run through the cases of formula (42):

(1) The case $✎(r,x) = \overline{✎(\cdot,\cdot)}$: suppose that, for each $x \in \theta_v(V_1)$, the following condition holds: $\forall r, s \in V_2(x) : ✎(r,x) = ✎(s,x) = \overline{✎(\cdot,\cdot)}$. In this case, we obtain for each $x \in \theta_v(V_1)$, the following result:

$$\sum_{r \in V_2} \beta_1\left(\theta_x(\dot{v})\right) \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)} = \beta_1\left(\theta_x(\dot{v})\right) \sum_{r \in V_2} \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)}$$
$$= \beta_1\left(\theta_x(\dot{v})\right). \tag{43}$$

In other words, if all authors of all texts contributing to the weight of a topic contribute to these texts according to the average activity, the weight of this topic in the ATN corresponds to that of the corresponding TTN. In this case, the average activity does not bias the weight of a topic in the ATN compared to the same topic in the corresponding TTN. Obviously, this scenario gives us a *neutral point* or, more specifically, a *calibration point* for the comparison of ATNs and TTNs. Such a calibration point allows us to interpret any down- or upward deviation of the topic weights in both networks, since no deviation means average activity and average number of active users. However, this consideration presupposes that $\omega_v = 1$ so that $\alpha_2 = \alpha_1 = \mathrm{id}$. If $\omega_v > 1$, then the number of authors of texts contributing to the weight of v is on average higher than expected on the basis of Wikipedia, so that the weight of the topic v in the ATN is "biased upwards" compared to the weight of the same topic in the corresponding TTN. Conversely, if $\omega_v < 1$, then the number of authors of texts contributing to the weight of v is on average smaller than expected, so that v's weight in the ATN is "biased downwards" compared to the weight of the same topic in the corresponding TTN. This scenario teaches us the different roles of α_2 and β_2 with respect to the weighting of the β_1 values: while β_2 operates as a function of the activities of authors, α_2 considers their number.

(2) The case $✎(r,x) < \overline{✎(\cdot,\cdot)}$: suppose for each $s \neq r$ that $✎(s,x) = \overline{✎(\cdot,\cdot)}$ while $✎(r,x) < \overline{✎(\cdot,\cdot)}$. Then, we conclude the following:

$$\beta_1\left(\theta_x(\dot{v})\right)\left(\sum_{t \in V_2 \setminus \{r\}} \frac{✎(t,x)}{\sum_{s \in V_2} ✎(s,x)} + \frac{1}{p} \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)}\right) < \beta_1\left(\theta_x(\dot{v})\right) \Longleftrightarrow$$
$$\sum_{t \in V_2 \setminus \{r\}} \frac{✎(t,x)}{\sum_{s \in V_2} ✎(s,x)} + \frac{1}{p} \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)} < 1 \Longleftrightarrow \tag{44}$$
$$\frac{1}{p} \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)} < \frac{✎(r,x)}{\sum_{s \in V_2} ✎(s,x)} \Longleftrightarrow 1 < p.$$

Thus, for $p > 1$, we penalize the contribution of a below-average active author of a text to the weight of the topic to which this text contributes. The different effects of $\omega_v \lessgtr 1$ have already been discussed.

(3) The case $\overline{\not\in}(r,x) \geq \overline{\not\in}(\cdot,\cdot)$: if we suppose now that $\forall s \neq r : \not\in(s,x) = \overline{\not\in}(\cdot,\cdot)$ while $\not\in(r,x) > \overline{\not\in}(\cdot,\cdot)$, we conclude that for $p > 1$, we reward the contribution of an above-average active author of a text to the weight of the topic to which this text contributes.

In a nutshell, α_2 and β_2 implement the following proportionality assumptions:

(i) By α_2 we penalize or reward under- or above-average coauthorships: the higher the above-average number of authors contributing to the texts of a topic, the higher the reward effect and the higher the weight of the topic. And vice versa, the lower the below-average number of authors contributing to the texts of a topic, the higher the penalty effect and the lower the weight of the topic.

(ii) By β_2 we penalize or reward under- or above-average activities of single authors: the higher the above-average activity of a single author contributing to a text of a topic, the higher the reward effect and the higher the contribution of this author-text pair to the weight of the topic. And vice versa, the lower the below-average activity of a single author contributing to a text of a topic, the higher the penalty effect and the lower the contribution of this author-text pair to the weight of the topic.

Finally, we define the functions γ_2 and δ_2 to get instantiations of the functions γ and δ of formula (11) of Definition 4 (or, in the generalized case, of formula (27) of Definition 7). This is done by means of the following auxiliary function:

$$\widetilde{\nu}_2(r,s) = \text{scale}\left(\nu_2(r,s), \overline{\nu}_2\right) \in \mathbb{R}^+, \tag{45}$$

where $\overline{\nu}_2$ estimates the average degree of coauthorship in Wikipedia according to formula (31). (We estimate $\overline{\nu}_2$ by means of 10,000 randomly selected Wikipedia articles so that $\overline{\nu}_2 \coloneqq 0.002,756.$) $\widetilde{\nu}_2(r,s)$ is a readjustment of $\nu_2(r,s)$ in relation to the mean value $\overline{\nu}_2$: the higher the above-average coauthorship, the higher the value of $\widetilde{\nu}_2$, and the lower the below-average coauthorship, the lower the value of $\widetilde{\nu}_2$. Then, we define

$$\begin{aligned}
\gamma &\coloneqq \gamma_2 = \text{id},\\
\delta\left[x \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{v}, y \underset{\theta^{\leftarrow}}{\overset{\theta}{\longleftrightarrow}} \dot{w}, r \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{v}, s \underset{\vartheta^{\leftarrow}}{\overset{\vartheta}{\longleftrightarrow}} \dot{w}, r \underset{\nu_{1.i}}{\overset{\nu_{i.1}}{\longleftrightarrow}} x, s \underset{\nu_{1.i}}{\overset{\nu_{i.1}}{\longleftrightarrow}} y, r \overset{\nu_i}{\longrightarrow} s, x \overset{\nu_1}{\longrightarrow} y\right] &\coloneqq\\
\delta_2\left(x \overset{\theta}{\longrightarrow} \dot{v}, y \overset{\theta}{\longrightarrow} \dot{w}, r \overset{\nu_{2.1}}{\longrightarrow} x, s \overset{\nu_{2.1}}{\longrightarrow} y, r \overset{\nu_2}{\longrightarrow} s, x \overset{\nu_1}{\longrightarrow} y\right) &=\\
\begin{cases}
\widetilde{\nu}_2(r,s) \cdot \beta_2\left(\theta_x(\dot{v})\right) \cdot \beta_2\left(\theta_y(\dot{w})\right), & (x,y) \in A_1 \wedge (r,s) \in A_2,\\
0, & \text{else.}
\end{cases}
\end{aligned} \tag{46}$$

In this definition, $\beta_2(\theta_x(\dot{v}))$ quantifies the link $x \overset{\theta}{\longrightarrow} \dot{v}$ and the link $r \overset{\nu_{2.1}}{\longrightarrow} x$ (cf. formula (11)), the product $\beta_2(\theta_x(\dot{v}))\beta_2(\theta_y(\dot{w}))$ quantifies the link $x \longrightarrow_{\nu_1} y$, and $\widetilde{\nu}_2(r,s)$ quantifies the link $r \longrightarrow_{\nu_2} s$. The calibration point of arc weighting is now reached under the conditions of the following scenario (for the first two conditions, see above):

$$\beta_2\left(\theta_x(\dot{v})\right) = \beta_1\left(\theta_x(\dot{v})\right), \tag{47}$$

$$\beta_2\left(\theta_y(\dot{w})\right) = \beta_1\left(\theta_y(\dot{w})\right), \tag{48}$$

$$\widetilde{\nu}_2(r,s) = 1. \tag{49}$$

Under these conditions, the authors r and s contribute to the texts x and y at an average level while interacting at an average level of coauthorship. In this case, the (co)authorship of both authors does not influence the strength of the corresponding arc in the ATN: in terms of neither reducing nor increasing $\nu_2(v,w)$. Note that the size of an ATN (i.e., the number of its arcs) is always less than or equal to that of the corresponding TTN, since the arcs present in a TTN are merely re-weighted in the corresponding ATN: no new arcs are added. The same holds for the order of the ATN since there is no node in a TTN for which there is no author authoring it.

Our instantiation of multiplex text and author topic networks has shown two points: firstly, we demonstrated a single-parameter setting as an element of a huge parameter space spanned by parameters such as p, $\overline{\nu}_2$, $\overline{\not\in}(\cdot,\cdot)$, $|\overline{\not\in}(\cdot,\cdot)|$, θ, α_1, α_2, β_1, β_2, γ_1, γ_2, δ_1, and δ_2. (In the latter eight cases, various information links are included as candidate parameters. Formula (42) shows, for example, that out of the six possible information links, only two are evaluated to instantiate β_2. Obviously, numerous alternatives exist to instantiate this function.) Secondly, anyone who complains about the apparently inherent parameter explosion in our approach should consider the hyperparameter spaces of neuronal networks as an object of parameter optimizations. Regardless of the heuristic character of our approach, compared to the black box character of neural networks, its settings are extensible on the basis of the schematic framework provided by Definition 8 of MTNs and the definitions it is based upon. At the same time, this approach guarantees interpretability as long as the different ingredients entering our model via formulas of the sort as formulas (26) and (26) fulfill this condition—in order to meet challenge P5.

3.2.4. Module 4: Network Randomization. Randomization is conducted to assess the significance of our findings. This is necessary because there is currently no related classification in the area examined here that can serve this role. To fill this gap, we compute the following randomizations:

(1) Baseline B1: a lower bound of a baseline is obtained by randomly assigning the object networks onto the gold standard (target) classes. This can be done by informing the assignment about the true cardinality of these classes (B11) or not (B12). We opt for B11 since this variant yields a higher *F*-score, making it more difficult to surpass. Of course, any serious network representation and classification model should go beyond this baseline. B1 will be averaged over 100,000 iterations.

(2) Baseline B2: an alternative is to randomize the input networks and to derive vector representations (according to Section 3.2.3), which ultimately undergo the same classification process as the original networks. That is, the input networks are randomly rewired to generate Erdős-Rényi (ER) graphs, for which we ask whether they are separable by the same classification model. (An alternative, not considered here, would be to randomize the topic classification of the underlying texts.) If this is successful in terms of high *F*-scores (the *F*-score is a measure of the accuracy of a classification, that is, the harmonic mean of its precision and recall), then we conclude that the network representation model or the operative classifier is not informative enough regarding the hypothetical class memberships of the input networks. Conversely, the lower the average *F*-scores obtained by classifying the randomized networks compared to the classification of the original ones, the more informative the representation model or the classification procedure regarding the underlying hypotheses. By keeping the model constant while varying the classifier, we can ultimately attribute this (non)informativity to the underlying representation model. Conversely, by keeping the classifier constant while varying the model, we can attribute this informativity to the classification model. B2 will be repeated 100 times.

(3) Baseline B3: a third baseline results from randomizing the matrices that form the input of the target classifiers. This means that instead of calculating graph invariants or similarity values to feed the classifiers, we use matrices whose dimensions are chosen uniformly at random from the domain of the corresponding invariants or (dis)similarity measures. (We require that the main diagonal of the random matrix is 1 and that it is symmetric.) If the classification based on the original networks does not exceed this baseline, we are again informed about a deficit of our representation model. Evidently, we are looking for models that significantly exceed this baseline; otherwise, we would have to accept that the same classifiers perform better on random values than on our feature model. B3 will be repeated 100 times.

(4) Baseline B4: finally, we start from randomly reorganizing the set of observations into random classes while using the same representation model to separate the resulting random gold standard. (Obviously, we have to prevent that the gold standard is ever part of the set of these randomizations.) We choose the variant of using randomized cardinalities of the random classes rather than keeping the sizes of the gold standard. Tests have shown that this approach tends to generate higher *F*-scores than the latter. If our network representation and classification model do not outperform this baseline, we learn that the underlying invariants used to characterize the networks are not *specific* enough; rather, they can be related to random classifications of the same objects using the same feature space. Obviously, we are looking for a model characterizing the gold standard (*tendency to specificity*) and not a random counterpart of it (*tendency to non-specificity*). B4 is averaged over 100 repetitions.

B1 is a lower bound: models that fall under this bound are obsolete. B2 concerns the evaluation of the network representation or classification model. B3 focuses on evaluating the classification model, and B4 aims at evaluating the specificity of the operative feature model.

3.2.5. Module 5: Network Quantification. Module 5 is a preparatory step for a subset of network similarity measures. This relates to so-called topology-based approaches to graph similarity [57, 93–96]. The idea behind this approach is to map input networks onto vectors of graph indices or invariants to compare them with each other. That is, graph similarity is traced back to similarity in vector space: the higher the number of indices for which two graphs resemble each other, the more similar the graphs. The apparatus that we employ in this context is described next.

3.2.6. Module 6: Graph Similarity Analysis. Our hypothesis about thematic networks on geographical places says that these networks are similar in terms of the skewness of their thematic focus and their network structure, regardless of whether the underlying texts are written by different communities and regardless of the framing theme. To test this hypothesis, we apply the framework of graph similarity measurement which allows for mapping the second of these three reference points by exploring the structure of topic networks as well as features of their nodes. Since graph similarity measurement is generally known to be computational complex, we take profit from the fact of dealing with *labeled* graphs. By using alignments of the labels of the nodes of the graphs to be compared, we reduce the time complexity of these approaches enormously.

The literature knows a number of approaches for graph similarity measurement. Among other things, this includes

the following approaches (see Emmert-Streib et al. [97] for an overview (cf. [98, 99]); the paper does not aim at a comprehensive study of them but focuses on a selected subset):

(1) *Graph Edit Distance-* (GED-) based approaches [100–102] and their relatives (e.g., the *Vertex and Edge* Overlap (VEO) [103])

(2) Spherical [90] or neighborhood-related approaches (cf. [99])

(3) Network topology-related approaches [57, 93–96, 103]

We will develop and test candidates of each of these classes.

GED-based methods are well studied in the area of web mining [104]. Since we are dealing with labeled graphs, we can compute the GED directly from the vertex and edge sets of the input graphs [99, 100]. Let $G_1 = (V_1, A_1, \mu_1, \nu_1, \lambda_1, \kappa_1)$ and $G_2 = (V_2, A_2, \mu_2, \nu_2, \lambda_2, \kappa_2)$ be two TNs, then their GED is computed as follows:

$$\begin{aligned}\mathrm{GED}(G_1, G_2) = &\left|V_1\right| + \left|V_2\right| - 2\left|V_{\mathscr{C}}(V_1) \cap V_{\mathscr{C}}(V_2)\right| \\ &+ \left|A_1\right| + \left|A_2\right| - 2\left|V_{\mathscr{C}}(A_1) \cap V_{\mathscr{C}}(A_2)\right| \in \mathbb{R}_0^+,\end{aligned} \tag{50}$$

where $V_{\mathscr{C}}(A_i) = \{(\dot{v}, \dot{w}) \mid (v, w) \in A_i\}, i = 1, \ldots, 2$. Since we are targeting graph similarities, we consider GES instead of GED, where overlaps of vertex and arc sets are equally weighted:

$$\mathrm{GES}(G_1, G_2) = 1 - \frac{1}{2}\left(\frac{\left|V_1\right| + \left|V_2\right| - 2\left|V_{\mathscr{C}}(V_1) \cap V_{\mathscr{C}}(V_2)\right|}{\left|V_1\right| + \left|V_2\right|} + \frac{\left|A_1\right| + \left|A_2\right| - 2\left|V_{\mathscr{C}}(A_1) \cap V_{\mathscr{C}}(A_2)\right|}{\left|A_1\right| + \left|A_2\right|}\right) \in [0, 1]. \tag{51}$$

The same is done in the case of Wallis' approach to graph distance [102], which is adapted as follows to get a similarity measure:

$$\mathrm{WAL}(G_1, G_2) = \frac{\left|V_{\mathscr{C}}(V_1) \cap V_{\mathscr{C}}(V_2)\right| + \left|V_{\mathscr{C}}(A_1) \cap V_{\mathscr{C}}(A_2)\right|}{\left|V_1\right| + \left|V_2\right| + \left|A_1\right| + \left|A_2\right| - \left|V_{\mathscr{C}}(V_1) \cap V_{\mathscr{C}}(V_2)\right| - \left|V_{\mathscr{C}}(A_1) \cap V_{\mathscr{C}}(A_2)\right|} \in [0, 1]. \tag{52}$$

A relative of GES is the *Vertex/Edge Overlap* (VEO) graph similarity measure [103]:

$$\mathrm{VEO}(G_1, G_2) = 2\frac{\left|V_{\mathscr{C}}(V_1) \cap V_{\mathscr{C}}(V_2)\right| + \left|V_{\mathscr{C}}(A_1) \cap V_{\mathscr{C}}(A_2)\right|}{\left|V_1\right| + \left|V_2\right| + \left|A_1\right| + \left|A_2\right|} \tag{53}$$

$$= 1 - \frac{\mathrm{GED}(G_1, G_2)}{\left|V_1\right| + \left|V_2\right| + \left|A_1\right| + \left|A_2\right|}, \quad \in [0, 1]. \tag{54}$$

Since node and arc weights are not taken into account by these measures, we compute the following variant of GES to close this gap:

$$\forall x, y \in \mathbb{R}_0^+ : \delta(x, y) = \frac{|x - y|}{\max(x, y)}, \quad \in [0, 1], \tag{55}$$

$$\forall v \in V_1 \forall w \in V_2 : \mathrm{wges}(v, w) = \begin{cases} \delta(\mu_1(v), \mu_2(w)), & \dot{v} = \dot{w}, \\ 0, & \text{else}, \end{cases} \in [0, 1], \tag{56}$$

$$\forall a = (v, w) \in A_1, \forall b = (x, y) \in A_2 :$$

$$\mathrm{wges}(a, b) = \begin{cases} \delta(\nu_1(a), \nu_2(b)), & \dot{v} = \dot{x} \wedge \dot{w} = \dot{y}, \\ 0, & \text{else}, \end{cases} \in [0, 1], \tag{57}$$

$$\mathrm{wges}(V_1, V_2) = \frac{|V_1| + |V_2| - 2\sum_{v \in V_1, w \in V_2} \delta(\mu_1(v), \mu_2(w))}{|V_1| + |V_2|}, \quad \in [0,1], \tag{58}$$

$$\mathrm{wges}(A_1, A_2) = \frac{|A_1| + |A_2| - 2\sum_{a \in A_1, b \in A_2} \delta(\nu_1(a), \nu_2(b))}{|A_1| + |A_2|}, \quad \in [0,1], \tag{59}$$

$$\mathrm{wges}(G_1, G_2) = \frac{\mathrm{wges}(V_1, V_2) + \mathrm{wges}(A_1, A_2)}{2}, \quad \in [0,1], \tag{60}$$

wges is sensitive to arc [99] *and* to vertex weights of TNs, the latter measuring the membership degree of the underlying texts to the topic represented by the corresponding vertex. We say that such measures are *dual weight-dependent*. These measures are of high interest since they cover more information on the underlying networks than single weight-dependent or even weight-independent measures (cf. the axiom of edge weight sensitivity of Koutra et al. [99]).

GED and its relatives share a view of similarity, according to which graphs are considered to be more similar the more (equally weighted) vertices and arcs they share. This notion of similarity is contrasted by spherical approaches (see above) as exemplified by DeltaCon [99]. Roughly speaking, according to DeltaCon, the more similar two graphs resemble each other from the perspective of their vertices, the more similar they are. Since DeltaCon is not dual weight-dependent, we consider a dual weight-dependent relative of it. To this end, we compute the cosine of the vectors of geodetic distances for each pair of equally labeled vertices. Since topic networks can differ in their order, we first have to align their node sets to make them comparable—this is also needed because we aim for a dual weight-dependent measurement. The required alignment is addressed by means of the following auxiliary graphs G_{12} and G_{21}:

$$\begin{aligned}
&\forall i, j \in \{1,2\}, i \neq j : G_{ij} = \left(V_{ij}, A_i, \mu_{ij}, \nu_i, l_{ij}\right),\\
&V_{ij} = V_i \cup \left\{w \in V_j \mid \nexists v \in V_i : \dot{v} = \dot{w}\right\},\\
&\forall v \in V_{ij} : \mu_{ij}(v) = \begin{cases} \mu_i(v), & v \in V_i,\\ 0, & \text{else,} \end{cases}\\
&\forall v \in V_{ij} : l_{ij}(v) = \begin{cases} l_i(v), & v \in V_i,\\ l_j(v), & \text{else.} \end{cases}
\end{aligned} \tag{61}$$

G_{12} and G_{21} are needed to make G_1 and G_2 comparable whose symmetric difference $V_1 \Delta V_2$ can be nonempty while their vertex labeling functions share the same co-domain (since G_1 and G_2 belong to the same multiplex topic network according to Definition 8). Obviously, $|G_{12}| = |G_{21}|$ so that for each $v \in V_i, w \in V_{ij} \backslash V_i; i, j \in \{1,2\}, i \neq j$, there is no path from v to w in G_{ij}. Cases in which no such path exists are denoted by $v \nsim w$; otherwise, if such a path exists, we denote by $\mathrm{ged}_{ij}(v, w)$ the length of the shortest path, that is, the geodetic distance between v and w in G_{ij}. As we deal with graph similarities, we first transform the distance values into similarity values:

$$\begin{aligned}
&\forall v, w \in V_{ij} : \mathrm{gep}_{ij}^{[\omega,\iota]}(v, w)\\
&= \begin{cases} 1 - \left(\dfrac{\mathrm{ged}_{ij}^{[\omega,\iota]}(v, w)}{|V_{ij}|}\right), & v, w \in V_i,\\ 0, & \text{else,} \end{cases} \quad \in [0,1],
\end{aligned} \tag{62}$$

gep is short for geodetic *proximity*. With the denominator $|V_{ij}|$, we penalize situations in which there is no path between v and w, that is, $v \nsim w$. The parameter $\omega \in \{\mathrm{w}, \neg\mathrm{w}\}$ specifies, whether the geodetic distance $\mathrm{ged}_{ij}^{[\omega,\iota]}$ and the geodetic proximity $\mathrm{gep}_{ij}^{[\omega,\iota]}$ are computed for the weighted (w) or unweighted ($\neg$w) variant of G_{ij}. If $\omega = \mathrm{w}$, we assume that each arc weighting value is normalized by means of the nonzero maximum value assumed by the arc weighting function for this network (this means that a graph G_2, which is obtained from a graph G_1 by multiplying the weights of all arcs of G_1 by a factor $c > 0$, will be equal to G_1 in terms of the graph similarity measure to be introduced now (insensitivity to certain scalings)). $\iota \in \mathbb{R}_0^+$ specifies the maximum geodetic distance to be considered: beyond this value, nodes w are considered to be of maximum geodetic distance $|V_{ij}|$ to v—irrespective of their real distance. For $\iota \geq |V_{ij}|$, we have to compute all geodetic distances. For values of $\iota \ll |V_{ij}|$ (e.g., $\iota = 2$), we arrive at variants of gep_{ij} that are less time complex. We consider the variant $\iota = \infty$ so that we take all path-related information into account. Now, we calculate the dual weight-dependent cosine of G_1 and G_2 as follows:

$$\forall v \in V_{12} \forall w \in V_{21} : \cos[\omega, \iota](v, w) = \frac{\sum_{x \in V_{12}, y \in V_{21}, \dot{x} = \dot{y}} \text{gep}_{ij}^{[\omega,\iota]}(v, x) \text{gep}_{ij}^{[\omega,\iota]}(w, y)}{\sqrt{\sum_{u \in V_{12}} \text{gep}_{ij}^{[\omega,\iota]}(v, u)^2} \sqrt{\sum_{u \in V_{21}} \text{gep}_{ij}^{[\omega,\iota]}(w, u)^2}}, \quad \in [0, 1], \tag{63}$$

$$\cos_{\mathscr{A}}[\omega, \iota, \phi, \mathbb{L}](G_1, G_2) = \frac{\sum_{v \in V_{12}, w \in V_{21}, \dot{v} = \dot{w} \in \mathbb{L}} \phi(v, w) \cos[\omega, \iota](v, w)}{\sum_{v \in V_{12}, w \in V_{21}, \dot{v} = \dot{w} \in \mathbb{L}} \phi(v, w)}, \quad \in [0, 1], \tag{64}$$

$$\cos_{\mathscr{V}}(G_1, G_2) = \frac{\sum_{v \in V_{12}, w \in V_{21}, \dot{v} = \dot{w}} \mu_{12}(v) \mu_{21}(w)}{\sqrt{\sum_{v \in V_{12}} \mu_{12}(v)^2} \sqrt{\sum_{w \in V_{21}} \mu_{21}(w)^2}}, \quad \in [0, 1], \tag{65}$$

$$\cos_{\mathscr{A}\mathscr{V}}[\omega, \iota, \phi, \mathbb{L}](G_1, G_2) = \frac{\cos_{\mathscr{V}}(G_1, G_2) + \cos_{\mathscr{A}}[\omega, \iota, \phi](G_1, G_2)}{2}, \quad \in [0, 1], \tag{66}$$

$\cos[\omega, \iota, \phi, \mathbb{L}](G_1, G_2)$ is the weighted cosine of the vectors of geodetic proximities of the same-named vertices in G_{12} and G_{21}. In this article, we consider two instantiations of parameter ϕ:

$$\forall v \in V_{12}, w \in V_{21}, \quad \dot{v} = \dot{w} : \phi_1(v, w) = 1, \tag{67}$$

$$\forall v \in V_{12}, w \in V_{21}, \quad \dot{v} = \dot{w} : \phi_2(v, w) = \max(d(v), d(w)), \tag{68}$$

where ϕ_1 implements an arithmetic mean. ϕ_2 is a function of the degree centrality [105] of its arguments: the more linked a topic in a network, the higher its impact onto the similarity of the input networks. The similarity view behind this approach is that while $\cos_X[\omega, \iota, \phi_1, \mathbb{L}]$, $X \in \{\mathscr{A}, \mathscr{A}\mathscr{V}\}$, treats all—peripheral or central—nodes equally, $\cos_X[\omega, \iota, \phi_2, \mathbb{L}]$ gives central nodes more influence. Take the example of two city networks [106]: it is plausible to say that if city networks look similar from the point of view of their central places, this should have more impact on the general similarity assessment than similarities from the point of view of peripheral locations. An extension would be to use more informative node weighting measures (e.g., closeness centrality). Finally, parameter $\mathbb{L}$ limits the number of vertices for which cosine values are computed. In the unlimited case, $\mathbb{L} \coloneqq \mathbb{L}_{12} = \{l_{12}(v) \mid v \in V_{12}\}$. It is easy to see that formulas (64)–(66) are similarity measures. For $X \in \{\mathscr{A}, \mathscr{V}, \mathscr{A}\mathscr{V}\}$, this can be shown as follows:

(1) *Symmetry*:
$\cos_X[\omega, \iota, \phi, \mathbb{L}](G_1, G_2) = \cos_X[\omega, \iota, \phi, \mathbb{L}](G_2, G_1)$
since formulas (63)–(66) are all symmetric.

(2) *Positivity:* since we are considering only positive arc weights, it always holds that

$$\cos_X[\omega, \iota, \phi, \mathbb{L}](G_1, G_1) \geq 0, \tag{69}$$

for any ω, ι, ϕ and $\mathbb{L} \neq \varnothing$.

(3) *Upper bound*: $\cos[\omega, \iota, \phi, \mathbb{L}](G_1, G_1) = 1$ for any ω, ι, ϕ and $\mathbb{L} \neq \varnothing$ and thus

$$\forall G_2 \neq G_1 : \cos[\omega, \iota, \phi, \mathbb{L}](G_1, G_1) \geq \cos[\omega, \iota, \phi, \mathbb{L}](G_1, G_2). \tag{70}$$

It is worth noticing that the range of values of formulas (63) and (65) is limited to $[0, 1]$, since the values of *gep* are always positive and we only consider positive membership values of texts to topic nodes.

So far we looked at measures that mostly processed the arc set A of TNs. This is contrasted by measures operating on topological indices of graphs. An example is NetSimile [107], which is based on the idea of characterizing networks by vectors of graph indices, which mostly draw on theories of social networks or egonets. Starting from seven local, node-related structural features (e.g., node degree, node clustering, or size of a node's egonet (see Berlingerio et al. [107] for the details of this approach)), it computes the mean and the first four moments of the corresponding distributions to generate 35-dimensional feature vectors per network where the Canberra Distance is used to compute their distances: let $\vec{x}, \vec{y} \in \mathbb{R}^k$ be two vectors, then their Canberra Distance is defined as

$$d_{\text{Can}}(\vec{x}, \vec{y}) = \sum_{i=1}^{k} \frac{\left|\vec{x}_i - \vec{y}_i\right|}{\left|\vec{x}_i + \vec{y}_i\right|}. \tag{71}$$

Soundarajan et al. [108] show that NetSimile is consistently close to the consensus among all measures studied by them, showing that it approximates the results of more complex competitors. This finding makes NetSimile a first choice in any comparative study of graph similarities.

Following on from this success, we introduce a topology-related approach to graph similarity, which draws on the hierarchical classification of the texts underlying the topic networks by reference to the *Dewey Decimal Classification* (DDC) (see Section 3.2.2). Starting from a pretest which essentially showed that graph invariants of complex network theory [109] do not sufficiently distinguish networks from their random counterparts, we decided to calculate a series of graph indices that evaluate the assignment of topics to the second level of the DDC. More specifically, we compute three node type-sensitive variants of the four cluster coefficient C_{ws} [110], C_{br} [111], C_{bbpv} [112], and C_{zh} [113] (cf. [114]). This variation can be exemplified by means of C_{ws}: to derive the desired variants from C_{ws}, we use the

following scheme, where mode $\in$ {intra, inter, heter} serves as a parameter to distinguish these alternatives (d_i is the degree of $v_i \in V$):

$$C_{\mathrm{ws}_{\mathrm{mode}}} = \frac{1}{n}\sum_{i=1}^{n} 2\frac{\mathrm{adj}_{\mathrm{mode}}(v_i)}{d_i^2 - d_i}, \quad \in [0,1], \tag{72}$$

where $\mathrm{adj}_{\mathrm{intra}}(v_i)$ is the number of adjacent neighbors of $v_i \in V$ sharing their 2nd level topic classification with v_i, $\mathrm{adj}_{\mathrm{inter}}(v_i)$ is the number of adjacent neighbors of v_i whose identical classification differs from that of v_i, and $\mathrm{adj}_{\mathrm{heter}}(v_i)$ is the number of adjacent neighbors of v_i whose classification differs among each other and from that of v_i (a 4th case is that v_i shares with a single neighbor its 2nd level topic while differing from the topics of all other neighbors). In this way, we compute for each of the cluster values C_{ws} (unweighted), C_{br} (unweighted), C_{bbpv} (weighted), and C_{zh} (weighted) three variants considering intra- and interrelational as well as heterogeneous type-sensitive clustering so that topic networks are finally represented by 12-dimensional feature vectors which are compared using the cosine measure. We call this approach *ToSi* (as short for *topological similarity*).

As a result of this candidate show of graph similarity measures, we consider the set of measures displayed in Table 5 for measuring the similarities of topic networks in order to shed light on Hypothesis 1, part (2).

3.2.7. Modules 7 and 8: Machine Learning and Classification Analysis. We conduct experiments in supervised learning with the aim of training classifiers to detect the layer (TTN or ATN) to which a topic network of a MTN belongs and the genre of the corpus from which the underlying LMN is derived. That is, our machine learning starts from a set of n genres $\mathcal{G}_i$, $i = 1, \ldots, n$, each of which is represented by a set $\mathbb{C}_i = \{C_{ij} \mid j = 1, \ldots, n_i\}$ of text corpora C_{ij} (see Figure 16). The set $\{\mathbb{C}_i \mid i = 1, \ldots, n\}$ defines a gold standard for which we assume that $\forall i, j = 1, \ldots, n,\ i \neq j : \mathbb{C}_i \cap \mathbb{C}_j = \varnothing$. Next, for each corpus C_{ij} of each genre $\mathcal{G}_i$, we span an LMN $\mathscr{L}(C_{ij}, 2)$ that in turn is used to derive a two-layer MTN $\mathscr{M}(C_{ij}, 2) = (\mathbb{M}_{ij}, \mathbb{D}_{ij})C_{ij}$ such that $\mathbb{M}_{ij} = \{M_{ij}, N_{ij}\}$ consists of exactly two topic networks: a TTN M_{ij} and an ATN N_{ij} both derived from $\mathscr{L}(C_{ij}, 2)$. In this way, we obtain the set $\mathbb{M}_{\mathrm{ttn}}$ and the set $\mathbb{M}_{\mathrm{atn}}$ of all TTNs and ATNs, respectively, both derived from $\mathscr{L}(C_{ij}, 2)$ according to Section 3.2.3. Next, each of the sets $\mathbb{M}_{\mathrm{ttn}}$ and $\mathbb{M}_{\mathrm{atn}}$ is randomized according to the procedure described in Section 3.2.4 (Baseline B2). In this way, we obtain the sets $\mathbb{M}'_{\mathrm{ttn}}$ and $\mathbb{M}'_{atn}$ as the randomized counterparts of $\mathbb{M}_{ttn}$ and $\mathbb{M}_{atn}$. As a result, we distinguish a range of classification experiments (1–14) only a subset of which will be conducted in Section 4 to tackle Hypothesis 1. We start with distinguishing TTNs from ATNs. The underlying classification hypothesis is as follows.

Hypothesis 2. Topic networks of the same layer (also called mode) (i.e., TTN or ATN) are more similar than networks of different modes (this concerns Scenario 1 (observed data) and Scenario 6 (randomized data) in Figure 16).

The similarity of TNs will be quantified by means of the apparatus of Section 3.2.6. Regardless of which genre (*urban* vs. *regional* vs. *encyclopedic communication*) the underlying corpus belongs to, Hypothesis 2 assumes that one can always distinguish TTNs from ATNs by their structure, while TTNs and ATNs are less distinguishable among themselves. This scenario is depicted in Figure 14 by Arrow 1. If we falsify the alternative to this hypothesis, we can assume that (poor, rich, or moderate) thematic intertextuality, as manifested by TTNs, is different from coauthorship-based networking of topics in ATNs. Collaboration- and intertextuality-based networking would then differ in a way that characterizes their layer. In order to test genre sensitivity as disregarded by Hypothesis 2, we carry out two experiments: one in which we classify TTNs (ATNs) by genre and one in which we combine both classifications by simultaneously classifying by genre and layer. When classifying by genre, we distinguish TNs derived from city wikis (*urban communication*), regional wikis (*regional communication*), and subnetworks of Wikipedia (*knowledge communication*) (see Section 3.2.2). Finally, we generate two control classes of wikis and Wikipedia-based networks outside of these three genres. The corresponding wikis are sampled in a way that their members are rather dissimilar. Our similarity measurement should therefore not work with them. In a nutshell, the underlying classification hypothesis is as follows.

Hypothesis 3. Topic networks of the same genre are more similar than those of different genres (this concerns Scenarios 2–4 (observed) and Scenarios 7–9 (random data) in Figure 16).

As we consider the genre-sensitive classification in the context of the layer-sensitive one, we get different classification scenarios:

(1) Scenario 2 in Figure 16 denotes the task of training a classifier that detects TTNs of the same genre while distinguishing TTNs of different ones. If this is successful, we can assume that the TTNs analyzed here are genre-sensitive or that the communication functions that we hypothetically associate with these genres influence the structure of these TTNs.

(2) Scenario 3 from Figure 16 regards the analog experiment for the genre-sensitive classification of ATNs.

(3) Scenario 4 concerns the alternative in which the modal difference of TTNs and ATNs is ignored in order to classify topic networks independently of their modal difference according to their underlying genre.

(4) This scenario is contrasted with Scenario 5, which considers classifiers for simultaneously detecting the genre and layer of TNs. The underlying classification hypothesis is as follows.

Hypothesis 4. Topic networks of the same layer and genre are more similar than networks of different layers or genres (this concerns Scenario 5 (observed data) and Scenario 10 (random data) in Figure 16).

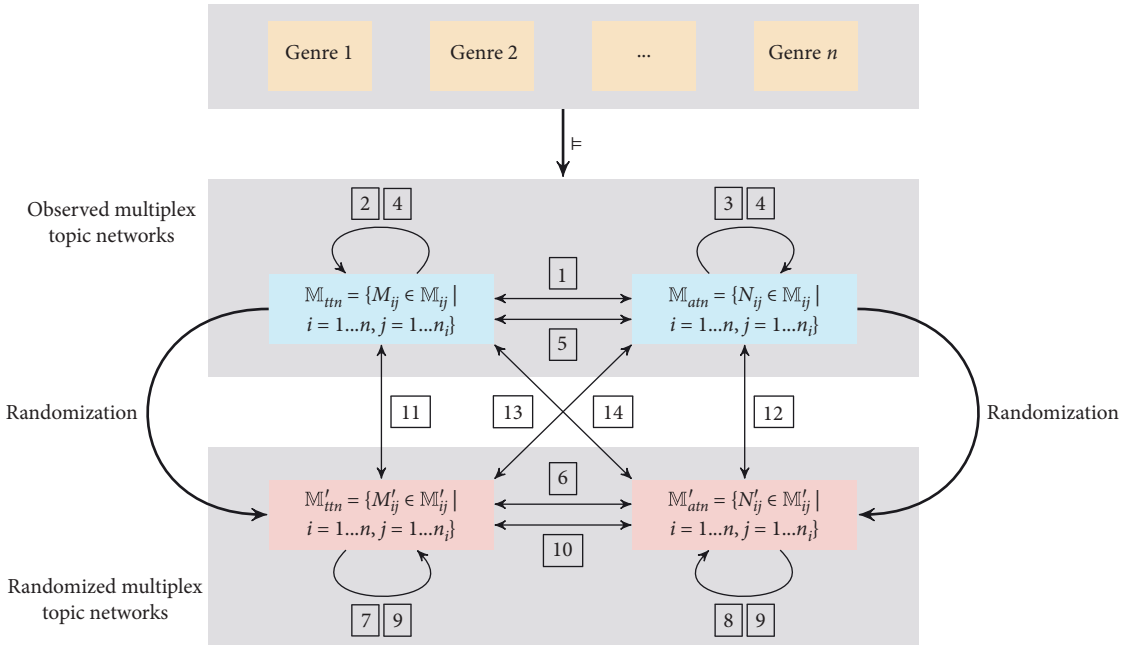


FIGURE 16: From sets of corpora of different genres to multiplex topic networks and their randomizations: corpora of different genres are the starting point for spanning LMNs which are then used to derive two-layer multiplex topic networks (⊨). In a second step, randomized counterparts according to Section 3.2.4 are derived from these MTNs to obtain a further basis for evaluating their significance. In this way, we arrive at fourteen candidate scenarios for classifying topic networks.

Falsifying the alternative to part (2) of Hypothesis 1 implies that TNs derived from corpora written by different communities by addressing different thematic frames (e.g., cities) appear nevertheless similar in their gestalt. Such a finding is very unlikely in cases in which the underlying corpora serve very different communication functions: Hypothesis 1 is not saying that everything is similar irrespective of the heterogeneity of the underlying function or the thematic orientation. Thus, a genre-oriented classification that shows that TNs of the same genre (serving a certain communication function and having a certain thematic orientation) are more similar than those belonging to different genres would rather correspond to such a finding. From this point of view, Hypotheses 3 and 4 are of interest: to deal with them experimentally could pave the way for testing the second part (2) of Hypothesis 1.

As explained in Section 3.2.4, we randomize input networks so that we obtain five additional classification scenarios labeled 6–10 in Figure 16. The experiments corresponding to these scenarios will be conducted here, as far as they concern the baseline scenario B2 of Section 3.2.4. Furthermore, scenarios are to be enumerated which attempt to distinguish observed networks directly from their randomized counterparts. In this context, Scenario 11 aims at distinguishing TTNs from their randomized counterparts by means of the classifiers trained to detect TTNs. Analogously, Scenario 12 considers ATNs in relation to their randomized counterparts, while Scenario 13 aims to separate observed topic networks (whether ATNs or TTNs) from randomized ones. Finally, Scenario 14 extends the latter scenario by trying to additionally account for the modal difference of ATNs and TTNs. These scenarios are only listed for theoretical reasons.

4. Experimentation

To test Hypothesis 1 and its relatives (i.e., Hypotheses 2–4), we conduct several experiments using two resources: a corpus of special wikis, called the *Frankfurt Regional Wiki Corpus*, and a corpus of subnetworks of Wikipedia that mostly contain information about cities and regions.

4.1. Tools and Resources. The *Frankfurt Regional Wiki Corpus* (FRWC) contains 43 wikis collected from online wiki lists (e.g., https://de.wikipedia.org/wiki/Regiowiki). Table 1 shows the statistics of this corpus, which is divided into three genres: CITIES relates to wikis describing certain cities, REGIONS includes wikis focusing on a specific region, while the residual class OTHERS collects wikis that are not off-topic w.r.t. regional communication but are unusual in their structure or the described rhemes. We consider only articles that are not redirects. Wiki authors use redirect pages to lead readers of articles with outdated, incorrect, or alternative spelling titles to the desired target page. We remove all such redirects and rewire all affected links accordingly. As a result, the number of processed articles is smaller than their overall number (see Table 1). In addition to the FRWC, we extracted a corpus of Wikipedia subgraphs (see Section 3.2.2 for the formal definition of these graphs and Table 3 for the corpus statistics). Subsequently, we denote the two variants in this Wikipedia corpus WP-REGIO-1 and WP-REGIO-2. We choose 25 articles

about cities or regions matching the titles of the wikis in the FRWC and additionally include the subgraphs of six off-topic articles to build two additional corpora, called WP-Others-1 and WP-Others-2, for purposes of comparison.

We process the content, link structure, and metadata (e.g., authorship-related information) of all articles in our corpora. This includes their history, that is, the chains of revisions which led to their current state. We do not consider past states of link structure and content itself but incorporate the authorship and the amount of content being added or removed per revision (see Section 3.2.3). The wikis considered here are based on MediaWiki. The structure of their articles varies from wiki to wiki so that HTML-based extractions are error-prone. To circumvent this problem, we use WikiDragon [115], a Java-based framework for importing and processing wikis offline.

For our experiments we used, adapted, and newly developed several tools including the so-called *GeneticClassifierWorkbench* (GCW), a Python library for performing feature selections and sensitivity analyses in classification experiments. Since our experiments are based on feature vectors with a size of sometimes more than 100 features, a complete sensitivity analysis of all feature combinations was not possible. Therefore, we conducted a genetic search for the best-performing subset of features due to maximizing the *F*-score. That is, a population of p features is evaluated and mutated over a number of t rounds. Instances which score best are saved unchanged for the next round and partly added in a slightly mutated form. The worst-performing instances are removed and replaced by random feature combinations. The Workbench is based on the Python library *scikit-learn* [116], allowing us to abstract from the underlying machine learning paradigm so that the same genetic search can be applied to optimize different classifiers. We experimented with neural networks which produced similar results on our test data but took too much time to be used for genetic searches and random baseline computations. Therefore, we decided for *Support Vector Machine*s (SVM) as the embedded method of supervised learning using the *Radial Basis Function* (RBF) as a kernel. Our source code is open source on GitHub (https://github.com/texttechnologylab/GeneticClassifierWorkbench).

4.2. Classification Experiments. We investigate the similarities of our seven corpora of regional wikis (Cities, Regions, and Others) and of Wikipedia-based subgraphs (WP-Regio-1, WP-Regio-2, WP-Others-1, and WP-Others-2) (each defining a corpus of texts) in order to test Hypothesis 1 and its derivatives, that is, Hypotheses 2–4. Thus, we distinguish up to seven target classes in our experiments. For reasons of simplicity, we call each element of these corpora *wiki* and each of the seven classes *genre*. Unless otherwise stated, the experiments are performed on all of them. In the case of WP-Regio-2 and WP-Others-2, we did not induce the corresponding ATNs, as some of these would have included several million edit events. Thus, in this case, we have at most five target classes. Each experiment includes three consecutive steps:

(1) *The all variant:* the first step, denoted by *all*, is a hyperplane parameter optimization and evaluation using the entire feature set. The optimized parameters of the respective classifier are then used in subsequent steps. Ideally, the parameters are optimized independently for each step, but this would have slowed down the genetic search.

(2) *The opt variant:* in the 2nd step, denoted by *opt*, genetic searches for optimal feature subsets are performed using a population of 20 feature vector instances and 50 rounds, trying to maximize the *F*-score of the classification. Note that these searches may only reach a local maximum.

(3) *The ext variant:* for experiments which are not conducted on random baseline data, we perform an extended genetic search for optimal feature subsets based on 20 instances and 500 rounds. In an additional step, a bit-wise genetic optimization attempts to further minimize the number of used features while keeping or even improving the *F*-score, using 20 instances and 500 rounds.

4.2.1. Graph-Similarity-Based Classification. Using the apparatus of Section 3.2.6, each TN (ATN or TTN) of each MTN is represented by a vector of values indicating its similarities to the wikis of the underlying experiment. Any such vector is separately computed for each of the 11 similarity measures of Table 5. Thus, if $\mathbb{T}$ is the set of all TNs of whatever mode (ATN or TTN) and genre (Cities, Regions, etc.) and if $\mathbb{T}' \subseteq \mathbb{T}$ is a subset of these TNs used in a classification experiment concerning the genres (target classes) $\text{Genre}\,i_1, \ldots, \text{Genre}\,i_j$ (cf. Figure 16), then each topic network $T \in \mathbb{T}'$ is represented for each similarity measure by a $|\mathbb{T}'|$-dimensional feature vector which is processed by the three-step algorithm described above. If for a given similarity measure the topic networks derived from wikis of the same genre are mapped to neighboring similarity vectors, then they belong to overlapping neighborhoods in vector space: *related networks are similar in their similarity and dissimilarity relations.* In this way, TNs of the same genre should become as recognizable as TNs of different genres. Now we see why a genetic search for optimal subsets of features is necessary: the reason is that otherwise we would assume that all dimensions of our feature vectors are equally informative—an assumption that is probably wrong.

Relating to Hypothesis 3, Tables 6 and 7 summarize our findings regarding the genre-sensitive classification of TTNs and ATNs, respectively. Cosine-based measures always perform best. Especially in the case of ATNs we see that accounting for arcs *and* for nodes secures better performance: dual weight-dependent measures (see Section 3.2.6) outperform single weight-dependent or weight-insensitive measures. However, in the case of TTNs, we also see that as long as we do not perform an extended optimization (ext), the measure $\cos_{\mathcal{AV}}[\neg w, \infty, \phi_1, L_{12}]$, which disregards arc weights, is a best performer. Of special interest is

Table 1: Statistics of the FRWC showing the number of articles with (#art. 1) and without (#art. 2) redirects, the number of revisions (#rev.), and the number of distinct authors (#authors).

Wiki	#art.1	#art.2	#rev.	#authors
Baden-Baden	999	844	3,576	138
Boppard	24	23	107	17
Cuxhaven	2,884	2,722	28,284	619
Dresden	11,479	9,796	76,776	2,702
Erfurt	2,275	2,267	30,314	129
Esslingen	252	219	2,646	353
Fürth	9,686	8,055	109,467	2,546
Görlitz	1,897	1,735	11,412	555
Hamm	16,602	14,439	99,307	1,353
Karlsruhe	38,870	25,575	306,143	11,002
Köln	3,925	3,184	13,394	400
Linz	6,776	4,250	28,923	343
Lüneburg	105	96	422	108
Lustenau	812	553	3,185	241
München	20,344	15,829	111,681	8,016
Münster	4,096	3,703	24,226	984
Olsberg	376	360	2,403	140
Reutlingen	583	545	3,122	368
Schiltach	505	489	560	14
Schorndorf	1,035	1,005	4,778	73
Strausberg	3,906	3,668	12,860	111
Stuttgart	1,260	1,076	6,784	228
Tübingen	4,749	4,211	38,540	1,513
Weißenburg	436	393	5,436	63
Wulfen	746	722	23,218	767
Würzburg	22,432	17,661	283,773	2,726

Wiki	#art.1	#art.2	#rev.	#authors
Ahrweiler	24,194	22,814	149,345	690
Attersee/Attergau	922	813	17,944	53
Dithmarschen	2,155	1,712	29,981	185
Ennstal	12,774	11,936	76,721	135
Franken	5,511	4,510	78,371	887
Göttingen	8,695	7,755	36,393	488
Niederbayern	33,751	20,504	196,525	1,392
Pforzheim-Enz	14,763	12,821	67,604	3,213
Rhein-Main	5,276	2,801	17,290	40
Rhein-Neckar	12,241	10,413	62,830	2,807
Sachenanhalt	4,644	4,173	36,264	1,153
Waldviertel	266	264	1,906	124

Wiki	#art.1	#art.2	#rev.	#authors
Graz	10,226	9,436	35,490	32
RegioWikiAT	12,085	8,551	113,436	3,221
Wallis	3,174	3,149	18,054	86
Wetzikon	1,737	1,302	23,999	446
Wien-Geschichte	45,473	43,919	296,467	402

Note: the last three columns disregard redirecting articles. Left table: genre CITIES; upper right: genre REGIONS; lower right: genre OTHERS. The German, Austrian, and Swiss wikis were downloaded in early 2018.

$\cos_{\mathscr{AV}}[w, \infty, \phi_2, L_{12}]$, the best performer regarding the classification of ATNs (Table 7), which is not only arc and node sensitive but also weights nodes as a function of their degree centrality and therefore covers the highest amount of structural information among all candidates considered here. This measure is also a robust candidate working at a high level in both experiments (it is the 2nd best performer in the case of TTNs if being optimized by an extended genetic search). Thus, we conclude that spherical measures clearly outperform GED-related approaches and especially network-topology-based approaches (ToSi and NetSimile) which perform worst: the kind of information we seek is apparently ignored or "abstracted away" by the latter measures. However, NetSimile has at least a high optimization potential (see the column ext in Table 6)—a potential which is missing in the case of ToSi. In any event, none of the measures considered here is outperformed by our baselines. But in Table 6, we also see that B3 (opt) approaches ToSi (all); in Table 7, we make analog observations also by example of other measures. A serious problem concerns NetSimile in relation to Baseline B2 regarding the classification of ATNs (Table 7): the baseline surpasses the topology-related measure whether being optimized (opt) or not (all). The graph indices collected by NetSimile have obviously difficulties in making observed networks distinguishable from their random counterparts—at least in some of the cases considered

Table 2: Building blocks of topic networks (texts, topics, words, agents, etc.), their relations according to Figure 8, and candidate procedures for weighting the corresponding arcs (last column).

Source	Relation	Target	Candidate procedure
Text	$\xrightarrow{\theta}$	Topic	text2ddc [72]
Topic	$\xrightarrow{\theta^{-}}$	Text	text2ddc^{-1}
Text	$\xrightarrow{\nu_1}$	Text	Measures of sentence/text similarity, text embeddings [78]
Agent	$\xrightarrow{\vartheta}$	Topic	Topic models [59]
Topic	$\xrightarrow{\vartheta^{-}}$	Agent	Topic models [59]
Agent	$\xrightarrow{\nu_{2,1}}$	Text	Edit networks [74]
Text	$\xrightarrow{\nu_{1,2}}$	Agent	Edit networks [74]
Agent	$\xrightarrow{\nu_2}$	Agent	Coauthorship [74, 79]
Word	$\xrightarrow{\vartheta}$	Topic	text2ddc [72], topic models [59]
Topic	$\xrightarrow{\vartheta^{-}}$	Word	text2ddc^{-1}, topic models [59]
Word	$\xrightarrow{\nu_{3,1}}$	Text	fastText, topic models [59]
Text	$\xrightarrow{\nu_{1,3}}$	Word	fastText, topic models [59]
Word	$\xrightarrow{\nu_3}$	Word	Word embeddings [77, 80–82]
—	—	—	—

Table 3: Wikipedia-based corpora: number of content articles (#articles n), revisions (#revisions n), and authors (#authors n) of non-redirecting articles in WP-Regio-1 ($n = 1$) and WP-Regio-2 ($n = 2$) of the German Wikipedia dump from 201807-01 (subgraphs 1–25); the variable n codes the nth orbit (see formula (30)). Subgraphs 26–31 are used to generate the corpora WP-Others-1 and WP-Others-2.

	Seed article	#articles 1	#revisions 1	#authors 1	#articles 2	#revisions 2	#authors 2
1	Ahrweiler	90	66,217	16,772	11,413	5,602,327	930,621
2	Dithmarschen	210	156,862	38,180	30,386	10,006,785	1,506,634
3	Dresden	1,615	1,180,743	239,747	127,675	27,746,644	3,566,957
4	Erfurt	943	850,786	179,282	100,052	23,644,822	3,158,299
5	Fürth	504	598,687	130,445	77,663	19,481,686	2,657,440
6	Görlitz	790	468,641	99,606	62,896	17,305,177	2,431,331
7	Göttingen	922	786,663	170,082	93,726	22,448,816	2,995,497
8	Hamm	764	697,437	150,502	82,099	20,436,567	2,799,384
9	Karlsruhe	1,021	842,723	180,652	97,984	23,178,185	3,103,192
10	KöIn	1,485	1,090,676	223,801	122,446	26,851,098	3,483,785
11	Linz	816	602,346	130,520	79,376	20,188,052	2,792,374
12	Metropolregion Rhein-Neckar	296	157,356	37,960	23,250	8,608,771	1,388,939
13	München	1,421	1,077,626	216,774	120,725	26,727,317	3,472,725
14	Munster	1,139	894,916	193,090	103,436	24,330,809	3,251,427
15	Niederbayern	239	142,392	33,551	22,466	7,796,961	1,222,744
16	Rhein-Main-Gebiet	390	297,276	65,804	42,238	12,750,028	1,870,354
17	Sachsen-Anhalt	603	459,933	96,116	59,565	16,392,237	2,291,304
18	Schorndorf	362	226,153	51,264	32,562	11,738,799	1,746,169
19	Steirisches Ennstal	43	19,702	6,322	4,400	2,101,467	386,487
20	Strausberg	265	215,854	49,617	30,284	10,602,198	1,579,390
21	Stuttgart	1,317	1,089,313	215,788	123,906	26,648,581	3,403,376
22	Tübingen	623	385,288	85,266	54,525	15,884,637	2,265,358
23	Wetzikon	204	145,207	33,914	20,607	8,044,399	1,306,780
24	Wien	1,380	874,419	170,952	102,792	23,357,095	3,087,254
25	Würzburg	959	885,109	185,495	106,381	24,484,274	3216,674
26	Hydraulik	121	59,874	19,400	8,287	3,600,636	700,341
27	Integralrechnung	194	75,082	21,787	6,708	2,663,563	508,606
28	Kernkraftwerk	287	196,202	49,279	20,773	8,195,232	1,387,491
29	Neuronales Netze	85	27,878	9,750	3,739	1,488,680	332,714
30	Schlacht bei Waterloo	200	97,290	25,614	18,674	6,990,403	1,097,749
31	Zecken	112	58,582	16,350	7,500	3,896,913	734,269

here. B3 is also of interest with regard to the classification of ATNs, which achieves F-scores of up to 40% and thus makes representation models based on measures such as NetSimile, ToSi, and *wges* problematic candidates. The values of B4 opt are also remarkably high and can therefore be regarded as a challenge for the measures.

Figure 17 shows that the baselines B1, B3, and B4 are outperformed by the results obtained for TTNs. However, it

Table 4: Estimates of the average number of active authors per Wikipedia article ($\overline{|✎(\cdot,\cdot)|}$) and the average activity of authors per article ($\overline{✎(\cdot,\cdot)}$) differentiated for the complete set of articles in the German Wikipedia (downloaded at 2018-07-01) with and without redirect articles (numbers of articles in parentheses).

Corpus of articles	$\overline{\lvert✎(\cdot,\cdot)\rvert}$	$\overline{✎(\cdot,\cdot)}$
Without redirects (2,195,812)	27.34	234.52
With redirects (3,657,483)	17.07	226.61

also shows that feature optimization affects the random baselines. This is particularly evident in the case of B3, which is based on random matrices. This gain in F-score can be explained by random numbers that allow the target classes to be separated—at least to some extent. These features are then selected by the genetic feature selection. The baseline results for ATNs show a similar picture (see Figure 17(b)). Regarding B2, we make the following observations in Figure 17(b) (for reasons of complexity, we did not consider all measures to compute B2): although the best B2 candidates are better than the average F-scores calculated on the basis of real data, B2 is clearly surpassed on average. Thus, we come to the conclusion that we found effective measures for comparing networks—this concerns in particular the spherical approach based on the cosine measure. From these experiments, we conclude the following:

(1) Hypothesis 3 is not falsified: we know the genre of a topic network by its structure. Note that this only concerns Scenarios 2 and 3 of Figure 16—Scenario 4 is not computed here. Similarly, by calculating our baselines, this also involves Scenarios 7 and 8 while ignoring Scenario 9. The classification benefits especially from information that is explored by dual weight-dependent measures. This holds regardless of the mode (ATN or TTN).

(2) Spherical measures should be preferred to GED-based measures and these in turn to topology-based measures:

$$\text{spherical} \succ \text{GED} \succ \text{topological}. \tag{73}$$

The boxplots in Figure 18 give another perspective on the classification results by summarizing the distributions of precision and recall values generated by the graph similarity measures. Except for the results on ATN using all features, the average precision is higher than the average recall. The figure also demonstrates the strong effect of feature selection.

So far, we considered classifications as a whole and thus abstracted from the scores obtained for individual genres. The boxplots in Figure 19 give insights into these genre-related scores regarding the classification of TTNs by means of the extended feature optimization (ext). The members of the genre Cities are well identified: in terms of recall and precision. The genre Regions is far less separable and causes many classification errors (low recall). Apparently, this class contains more heterogeneous TTNs. In any event, the Wikipedia-based genres WP-Regio-1 and WP-Regio-2 are very well separated. By contrast, instances of the category Others are extremely difficult to detect (as predicted in Section 3.2.7). Similarly, elements of the classes WP-Others-1 and WP-Others-2 are difficult to identify—albeit to a minor degree. Thus, we conclude that the upper bound of separability concerns Wikipedia-based regional wikis. The corresponding subgraphs are very similar. This upper bound is approached by city wikis. Region wikis are less homogeneous, making the corresponding class Regions rather blurred and therefore question its status as a genre. Figure 20 shows the corresponding results of classifying ATNs. The general picture is quite similar to that of the TTNs.

We take another perspective on the results to examine classification errors. The best results on TTNs using all features is achieved by $\cos_{\mathscr{MV}}[\neg w, \infty, \phi_1, \mathbb{L}_{12}]$. Figure 21 shows to what degree wikis of a target class are wrongly classified using this measure. The labels show the proportion of the categories according to the gold standard (top) and the classification result (bottom). The picture is diverse, but some details become clear: wikis of the classes Regions and Others are often falsely categorized as Cities. City wikis on the other hand are wrongly classified as WP-Others-1 or WP-Regio-1.

Genetic feature selection has proven to increase F-score significantly. In the extended optimization (ext), the last step is to minimize the number of features used. Since our features stand for similarities to networks, we have to ask whether some of the wikis underlying these networks are more relevant for the differentiation of the target classes than others—possibly because of their prototypical status. If all wikis were equally important, an equal distribution of the frequencies with which these features are selected by the genetic optimization would be expected. Figure 22 shows the corresponding rank frequency distribution: it shows that we are far from evenly distributed features. From this, we conclude that the selection of features is indispensable and that the underlying wikis are very different in their roles in our classification experiments.

Next, we try to distinguish TTNs from ATNs thereby addressing Hypothesis 2 (or more specifically, Scenario 1 of Figure 16). The error analysis in Figure 23 shows that networks of these two modes are not separable using our approach. Table 8 differentiates this outcome by reporting the results obtained for different measures. It shows that this classification scenario is far exceeded by Baseline B1 and is therefore irrelevant. From this result, we conclude that ATNs are so similar to their corresponding TTNs that they cannot be distinguished by our measures, or alternatively, our similarity measures are not suitable to distinguish them. This is not surprising, as the order and the size of an ATN always correspond to the order and the size of the TTN from which it was derived, so that they can only differ by the weighting of their nodes and arcs. By concerning Hypothesis 4 and thus by distinguishing twelve target classes (in the case of WP-Others-2 and WP-Regio-2, we do not induce ATNs), Table 8 shows a somehow different scenario: though the F-scores are still rather low, Baseline B1 is clearly outperformed when using a cosine measure for graph similarity measurement. From this observation, we conclude that while Hypothesis 2 is

Table 5: The list of measures of graph similarity used for computing the similarities of topic networks.

	Measure	Approach	Formula	Reference
1	GES	Graph edit similarity	(51)	[100]
2	WAL	Graph edit similarity	(52)	[102]
3	VEO	Vertex and edge overlap	(54)	[103]
4	wges	Weighted graph edit similarity	(60)	
5	$\cos_{\mathscr{A}}[w, \infty, \phi_1, L_{12}]$	Cosine graph similarity	(64)	
6	$\cos_{\mathscr{A}\mathscr{V}}[w, \infty, \phi_1, L_{12}]$	Cosine graph similarity	(66)	
7	$\cos_{\mathscr{A}\mathscr{V}}[w, \infty, \phi_2, L_{12}]$	Cosine graph similarity	(66)	
8	$\cos_{\mathscr{A}}[\neg w, \infty, \phi_1, L_{12}]$	Cosine graph similarity	(64)	
9	$\cos_{\mathscr{A}\mathscr{V}}[\neg w, \infty, \phi_1, L_{12}]$	Cosine graph similarity	(66)	
10	NetSimile	Topological similarity	(70)	[107]
11	ToSi	Topological similarity	(72)	

Table 6: F-scores of classifying TTNs into seven target classes (Cities, Regions, Others, WP-Regio-1, WP-Regio-2, WP-Others-1, and WP-Others-2) by means of SVMs using RBF kernels.

	Measure	all	opt	ext	B1	B3 all	B3 opt	B4 all	B4 opt
1	GES	0.653	0.753	0.798	0.143	0.130	0.286	0.121	0.213
2	WAL	0.649	0.751	0.788	0.143	0.130	0.286	0.109	0.216
3	VEO	0.677	0.773	0.816	0.143	0.130	0.286	0.120	0.221
4	*wges*	0.559	0.620	0.650	0.143	0.130	0.286	0.120	0.199
5	$\cos_{\mathscr{A}}[w, \infty, \phi_1, L_{12}]$	0.638	0.722	0.764	0.143	0.130	0.286	0.119	0.211
6	$\cos_{\mathscr{A}\mathscr{V}}[w, \infty, \phi_1, L_{12}]$	0.729	0.768	**0.853**	0.143	0.130	0.286	0.125	0.223
7	$\cos_{\mathscr{A}\mathscr{V}}[w, \infty, \phi_2, L_{12}]$	0.694	0.766	0.832	0.143	0.130	0.286	0.127	0.229
8	$\cos_{\mathscr{A}}[\neg w, \infty, \phi_2, L_{12}]$	0.642	0.681	0.717	0.143	0.130	0.286	0.122	0.212
9	$\cos_{\mathscr{A}\mathscr{V}}[\neg w, \infty, \phi_2, L_{12}]$	**0.742**	**0.773**	0.790	0.143	0.130	0.286	0.102	0.156
10	NetSimile	0.479	0.629	0.722	0.143	0.130	0.286	0.127	0.229
11	ToSi	0.390	0.433	0.465	0.143	0.130	0.286	0.108	0.229

Column "all": F-scores, if all features are used by the similarity measure (row). Column "opt": F-scores, if a subset of features selected by the genetic search is used. Column "ext": F-scores, if a subset of features selected by the extended genetic search is used. The last five columns display the F-scores of the random baselines B1, B3, and B4, in the case of B3 and B4 differentiated for the variants *all* and *opt*.

Table 7: F-scores of classifying ATNs into five classes (Cities, Regions, Others, WP-Regio-1, and WP-Others-1) by means of SVMs using RBF kernels.

	Measure	all	opt	ext	B1	B2 all	B2 opt	B3 all	B3 opt	B4 all	B4 opt
1	GES	0.598	0.649	0.752	0.200	0.226	0.325	0.182	0.397	0.176	0.294
2	WAL	0.610	0.635	0.707	0.200	0.168	0.222	0.182	0.397	0.158	0.289
3	VEO	0.636	0.706	0.783	0.200	0.213	0.306	0.182	0.397	0.170	0.308
4	*wges*	0.458	0.576	0.618	0.200	0.311	0.348	0.182	0.397	0.173	0.281
5	$\cos_{\mathscr{A}}[w, \infty, \phi_1, L_{12}]$	0.567	0.673	0.737	0.200	—	—	0.182	0.397	0.173	0.300
6	$\cos_{\mathscr{A}\mathscr{V}}[w, \infty, \phi_1, L_{12}]$	**0.740**	0.777	0.854	0.200	0.242	0.440	0.182	0.397	0.181	0.320
7	$\cos_{\mathscr{A}\mathscr{V}}[w, \infty, \phi_2, L_{12}]$	0.612	**0.816**	**0.875**	0.200	—	—	0.182	0.397	0.187	0.340
8	$\cos_{\mathscr{A}}[\neg w, \infty, \phi_2, L_{12}]$	0.559	0.600	0.652	0.200	—	—	0.182	0.397	0.182	0.307
9	$\cos_{\mathscr{A}\mathscr{V}}[\neg w, \infty, \phi_2, L_{12}]$	0.721	0.811	0.865	0.200	0.240	0.464	0.182	0.397	0.182	0.317
10	NetSimile	0.467	0.507	0.610	0.200	0.494	0.602	0.182	0.397	0.173	0.272
11	ToSi	0.431	0.567	0.585	0.200	-		0.182	0.397	0.179	0.254

Column "all": F-scores using all features in terms of the respective similarity measure. Column "opt": using a subset of features detected according to a genetic search. Column "ext": subset selection according to extended genetic optimization. Additionally, F-scores of random baselines B1, B2, B3, and B4 are displayed, in the latter three cases differentiated for the variants *all* and *opt*.

falsified, there is at least a potential regarding the simultaneous distinction of genre and mode: ATNs do not uniformly resemble their corresponding TTNs.

So far we considered part (2) of Hypothesis 1 by showing that TTNs (and also ATNs) with similar functions resemble each other, while differing from networks of other genres. It remains to be shown that these networks are also thematically focused—in a highly skewed manner. To test this, we fit power laws to the distributions of node weights in TTNs. Remember that these weights result from detecting textual instances of the topic represented by the respective node so that the more such instances are detected, the more salient the topic in the network. Fitting a power law to such a distribution means that there is a minority of topics or just one topic that surpasses all other topics in its importance, while the

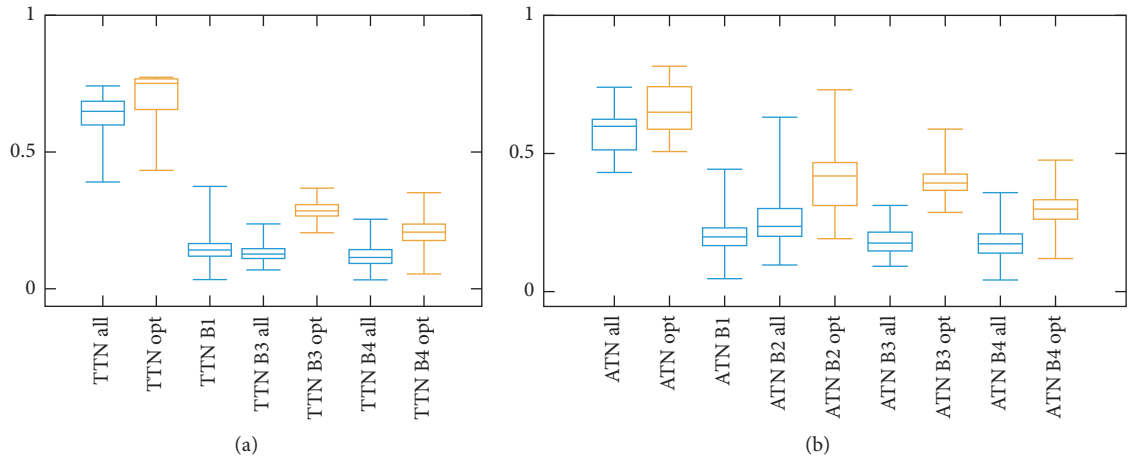


Figure 17: (a) Boxplots of F-scores obtained for classifying TTNs contrasted by the baselines B1, B3, and B4. (b) Boxplots of F-scores obtained for classifying ATNs contrasted by the baselines B1, B2, B3, and B4.

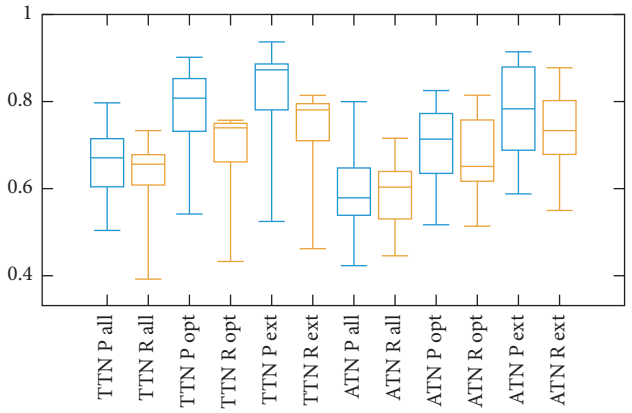


Figure 18: Boxplots of precision P and recall R values (y-axis) induced by the measures of Table 5 and underlying the F-scores of Table 6 (first six columns) and Table 7 (last six columns). Distributions are distinguished by considering all features (all) or subsets of them generated by the genetic optimization *opt* or *ext*.

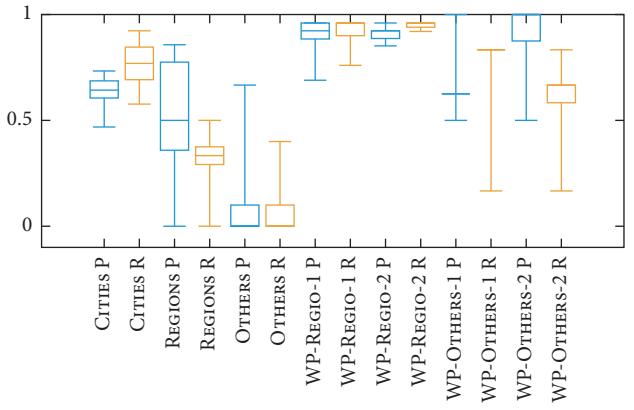


Figure 19: Boxplots of precision (P) and recall (R) values (y-axis) induced by the measures of Table 5 underlying the F-scores of Table 6. Distributions are distinguished by the respective target class of the classification.

majority of topics are of little or no importance. The boxplots in Figure 24(a) show the distribution of the exponents of the power laws fitted to these distributions, differentiated by the genres considered here. To assess the goodness of the fittings, we compute the adjusted R-squares and display the value distributions in Figure 24(b). Obviously, the fits are very good (the adjusted R-squares are on average above 95%) while the averages of the exponents range between 0.5 and 1.5: from this analysis, we conclude that the underlying wikis are all thematically focused and skewed by dealing with a minority of topics in depth. The five most detected DDC labels per genre are shown in Table 9. It shows that *Transportation; ground transportation* is by far the most dominant topic in city wikis and in region wikis. *Obviously, these wikis are thematically focused in a highly skewed manner.*

It remains to be shown that our findings about urban wikis neither depend on the distances of the

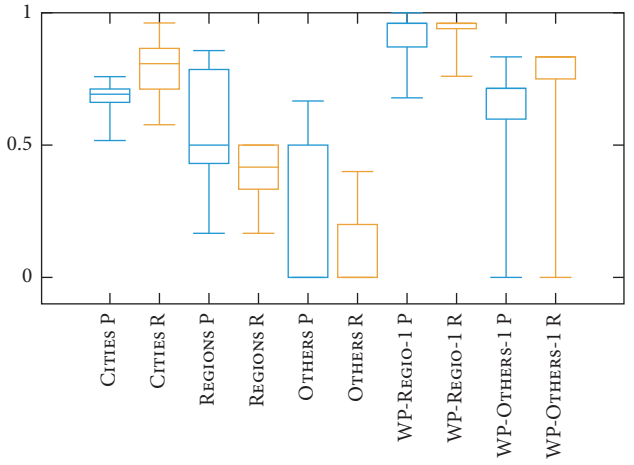


Figure 20: Boxplots of precision P and recall R values (y-axis) induced by the measures of Table 5 underlying the F-scores of Table 7. Distributions are distinguished by the respective target class of the classification.

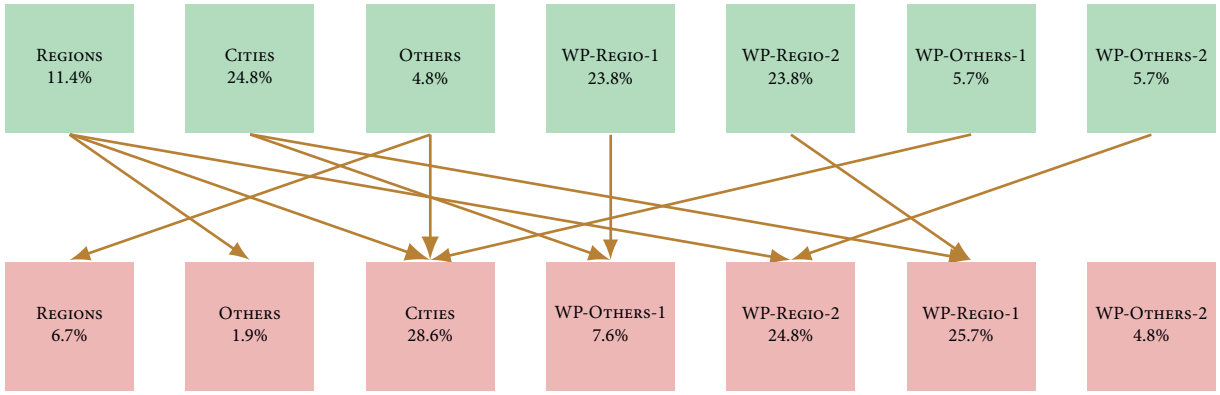


Figure 21: Error analysis regarding the classification of TTNs by means of $\cos_{\mathscr{AV}}[\neg w, \infty, \phi_1, \mathbb{L}_{12}]$.

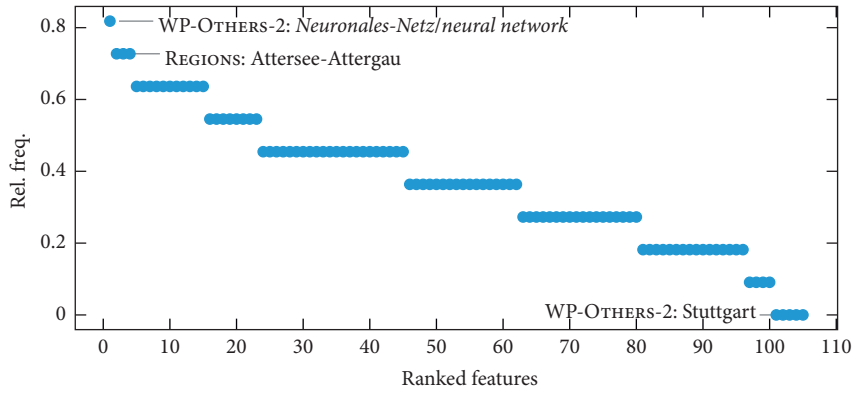


Figure 22: Ranking of the relative frequencies of features as a result of being selected by the extended genetic feature optimization in the classification of TTNs.

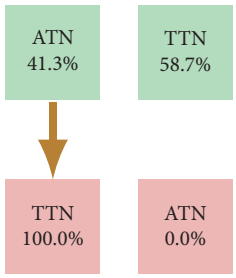


Figure 23: Error analysis regarding the classification of TTNs vs. ATNs by means of $\cos_{\mathscr{AV}}[\neg w, \infty, \phi_1, \mathbb{L}_{12}]$.

corresponding places nor on the communities writing these wikis. Figure 25 shows that the similarities detected by us do hardly correlate with the underlying distances of the places. In the heatmap in Figure 25(a), a connection between two city wikis is the greener, the closer, and the more similar they are to each other, while a pair of wikis is the more red, the less similar, and the more distant they are. Similarity is measured by $\cos[w, \infty, \phi_1, \mathbb{L}_{12}]$ while distance is converted into closeness and normalized to the unit interval (the values of the heatmap scale to $[-1, 1]$ by calculating $-1 +$ closeness + similarity). Figure 25(b) shows that there is hardly a tendency to being more similar when being more close to each other. The lower similarity values are mostly induced by the rather unusually small wikis such as Boppard (see Table 1). Figure 26 shows the Fuzzy Jaccard of the communities underlying the wikis, that is, the overlap of these communities weighted by the activities of their authors: the lower the number of shared authors of two wikis and the less active these authors, the lower the fuzzy overlap of these wikis. The Fuzzy Jaccard is computed as follows (cf. [117]): let $\text{authors}(\mathbb{W})$ be the set of all registered users contributing to any of the wikis in $\mathbb{W} = \text{Cities} \cup \text{Regions}$, $\text{Others} \cup \text{WP-Regio-1} \cup \text{WP-Others-1}$ and let $\text{texts}(W)$ be the set of all (nonredirect) articles of wiki $W \in \mathbb{W}$, then we compute

$$\forall A, B \in \mathbb{W} : J_\mu(A, B) = \frac{\sum_{r \in \text{authors}(\mathbb{W})} \mu_{A \cap B}(r)}{\sum_{r \in \text{authors}(\mathbb{W})} \mu_{A \cup B}(r)} \in [0, 1], \tag{74}$$

where

Table 8: Left: *F*-scores obtained for different measures and optimizations by classifying ATNs vs. TTNs according to Scenario 1 of Table 16—two target classes are considered. B1 considers Scenario 6 of Figure 16. Right: *F*-scores obtained for different measures and optimizations by classifying simultaneously for mode and genre according to Scenario 5—twelve target classes are considered. B1 considers Scenario 10.

	Measure	all	opt	ext	B1
1	GES	0.370	0.370	0.370	0.500
2	VEO	0.370	0.370	0.370	0.500
3	$\cos[w, \infty, \phi 1, L_{12}]$	0.370	0.370	0.370	0.500
4	$\cos[\neg w, \infty, \phi_1, L_{12}]$	0.370	0.370	0.370	0.500

	Measure	all	opt	ext	B1
1	GES	0.152	0.178	0.194	0.082
2	VEO	0.181	0.228	0.259	0.082
3	$\cos[w, \infty, \phi_1, L_{12}]$	0.315	0.363	0.407	0.082
4	$\cos[\neg w, \infty, \phi_1, L_{12}]$	0.284	0.339	0.409	0.082

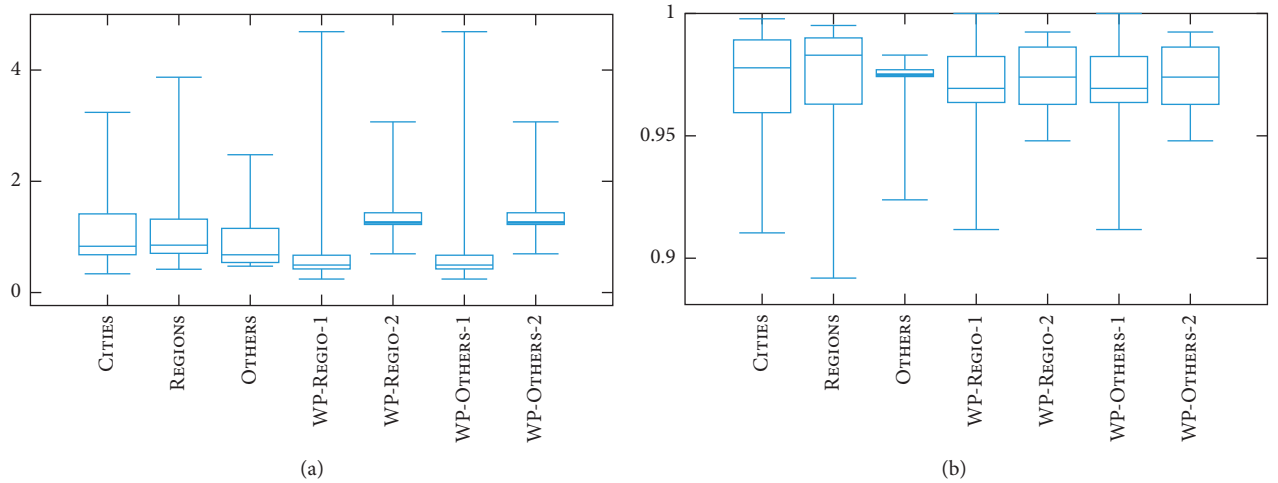
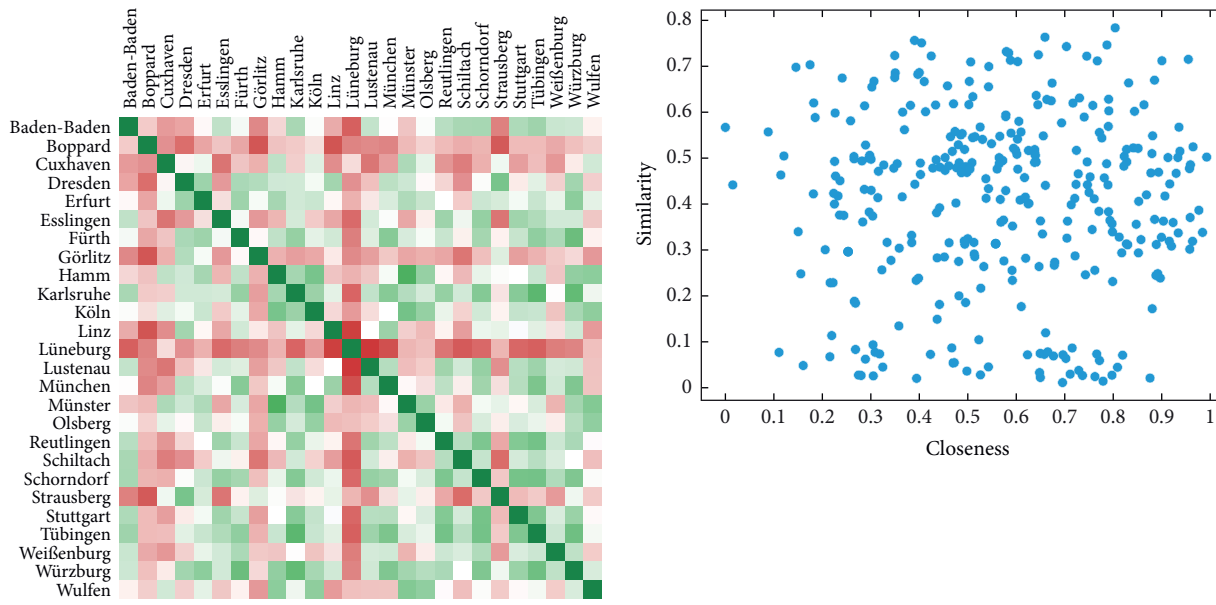


Figure 24: (a) Boxplots of the distribution of the exponents of the power laws fitted to the weight distributions of the nodes in the TTNs differentiated by the target classes. (b) The boxplots of the corresponding fitting accuracies computed by means of the adjusted *R*-squared.

Table 9: The five most detected DDC labels for the genres Cities, Regions, and Others.

Rank	Genre	Node weight sum	Avg weight	DDC	Description
1	City	10,325.830	397.147	388	Transportation; ground transportation
2	City	2,404.631	92.486	943	Central Europe; Germany
3	City	1,570.010	60.385	726	Buildings for religious purposes
4	City	1,512.536	58.174	725	Public structures
5	City	964.262	37.087	711	Area planning
1	Region	5,127.546	427.296	388	Transportation; ground transportation
2	Region	1,692.267	141.022	943	Central Europe; Germany
3	Region	1,385.013	115.418	726	Buildings for religious purposes
4	Region	1,289.722	107.477	551	Geology, hydrology & meteorology
5	Region	1,171.656	97.638	796	Athletic & outdoor sports & games
1	Other	5,335.555	1,067.111	929	Genealogy, names & insignia
2	Other	1,640.042	328.008	726	Buildings for religious purposes
3	Other	715.084	143.017	723	Architecture from ca. 300 to 1399
4	Other	701.298	140.260	725	Public structures
5	Other	680.309	136.062	720	Architecture



(a) (b)

FIGURE 25: (a) The heatmap of thematic similarity and spatial closeness among city wikis. Red means that the wikis are thematically dissimilar and distant in space; green means that they are thematically similar and close in space. (b): the distribution of the similarities (y-axis) as a function of the closenesses (x-axis) of the different pairs of city wikis.

$$\mu_{A\cap B}(r) = \min\left(\frac{\sum_{x\in \text{texts}(A)} \measuredangle(r,x)}{\sum_{s\in \text{authors}(A)}\sum_{x\in \text{texts}(A)} \measuredangle(s,x)}, \frac{\sum_{x\in \text{texts}(B)} \measuredangle(r,x)}{\sum_{s\in \text{authors}(B)}\sum_{x\in \text{texts}(B)} \measuredangle(s,x)}\right),$$
$$\mu_{A\cup B}(r) = \max\left(\frac{\sum_{x\in \text{texts}(A)} \measuredangle(r,x)}{\sum_{s\in \text{authors}(A)}\sum_{x\in \text{texts}(A)} \measuredangle(s,x)}, \frac{\sum_{x\in \text{texts}(B)} \measuredangle(r,x)}{\sum_{s\in \text{authors}(B)}\sum_{x\in \text{texts}(B)} \measuredangle(s,x)}\right). \quad (75)$$

Figure 25 shows that while among the Wikipedia-based extractions the overlap is remarkably high, it does nearly not exist between any of the city or region wikis: these wikis are written by mostly completely different communities. The picture is not different if one considers all authors—registered and unregistered.

5. Discussion

Section 4 has shown that topic networks, whether TTNs or ATNs, are similar if they belong to the same genre, while they are characterized by a high degree of thematic focusing. In order to operationalize this notion of network similarity, we tested further or newly developed 11 different measures of network similarity by relying on four different paradigms of measuring the similarity of graphs (see Table 5 and the discussion of graph/network similarity measures in Section 3.2.6) as instantiated by the complex networks studied here. All these measures and paradigms come along with a different notion of network similarity. We have shown that a subclass of them, especially cosine-based measures of network similarity, allow for detecting similarities of topic networks in line with Hypotheses 3 and 4. At the same time, the concept of network similarity underlying this class of dual weight-dependent measures seems to be the most promising from a research point of view, as it is based on node and arc weights and instantiates a very intuitive concept of network similarity: The more similar the two networks are from the perspective of the more of their nodes, the more similar they are. Thus, at the level of thematic abstraction examined here, there seems to be a hidden tendency to write about very prominent topics when it comes to thematizing places and linking the underlying texts in such a way that the resulting networks become almost indistinguishable.

Starting from this kind of thematic distortion of VGI as conveyed by online media, we now ask for a more general explanation of our findings. The candidate we are considering for this purpose is given by *Cognitive Maps* (CM), which were introduced as models of the cognitive representation and processing of spatial information to explain a number of different cognitive biases. Because of bridging the gap between geographical information and its biased representation, CMs promise to be a candidate for our task. At the same time, this notion allows for the connection of cognitive geography on the one hand and our generalized model of linguistic encoding of geographical information on the other (see Figure 1). The reason is that as mental representations, CMs are seen to integrate a wide range of representations of spatial objects, their relations, and

Figure 26: The Fuzzy Jaccard overlap of the communities of registered authors of the wikis in the corpora Cities, Regions, Others, WP-Regio-1, and WP-Others-1 weighted by means of the writing activities of the authors: the greener the link, the higher the fuzzy overlap.

thematic units (see below). We may argue now that we developed a method to represent and analyze a particular type of thematic information which can be subsumed under the latter list. If this is true, then the thematic distortion observed by us could be seen as a result of the biased processing of geographic information by a community of agents dealing with the same place to generate a common cognitive map, thereby manifesting a particular type of distributed cognition. When creating such a common CM of the same place, agents tend to focus on a highly selected set of rhemes (see Figure 1), even if there is no explicit agreement among these agents about this selection and even if there is little or no direct communication between them and also irrespective of the focal place. It seems that the agents participate in processes of distributed cognition in such a way that their own thematically distorted maps flow into the formation of a shared, stable but likewise distorted "thematic map." These maps then appear as the result of a sort of swarm behavior regarding the formation of a particular distribution of the preference and salience of certain place-related rhemes. From this perspective, topic networks serve as models of these thematic maps which in turn are parts of CMs. To underpin this interpretation, we briefly summarize the research on CMs and, above all, ask about distortions that are distinguished by the research in this area.

Understood as mental representations of spatial knowledge, CMs have been subject of scientific work for

decades. Starting from different disciplinary perspectives, this research provides insights into how people perceive their environment, think about it, and how this influences their spatial behavior. The interdisciplinary research on CMs has led to a multitude of notions, research designs, and outcomes, the integration of which is still pending. Over the years, researchers worked, for example, with different terms for the mental representations in question such as *cognitive maps* [118], *environmental images* [119], *mental maps* [120], *mental sketch maps* [121], *narrative space maps* [122], or *internal representations* [123], where the constituent *map* is most common. However, there has been a discussion as to whether the term *map* is generally misleading. In this context, Kitchin ([124], 3 pp.) distinguishes approaches that understand CMs as

(1) Three-dimensional maps

(2) An analogy to maps (because of their map-like characteristics)

(3) A metaphor for maps (because they function as if they were maps), or

(4) A hypothetical construct used to explain spatial behavior

While we refer to cognitive maps as an auxiliary notion, we adhere to the fourth of these variants. Regardless of this discussion, there is a greater consensus on some characteristics of CMs as mental representations: CMs are understood as complexes of mental images and concepts that humans have in mind when thinking about places, their *location* (in terms of distance and direction), *accessibility* (regarding questions like how to get there), and the *meanings* associated with them. They serve as a means of understanding spatial circumstances and as a frame of reference for the interpretation, preference, and prediction of spatial structures, their relations, and events in which they participate (see [125], 100 pp, 313), ([120], 3), and ([119], 5p.). Beyond that, they also serve as a basis for decision-making regarding spatial behavior (e.g., in route planning). In a nutshell, humans activate, generate, and utilize CMs in spatial thinking and spatial behavior (cf. [126], 233). CMs are distinguished according to the entities they model. Kitchin and Blades ([127], 5p) distinguish CMs of *object spaces* (e.g., rooms and cars), *environmental spaces* (e.g., buildings, streets, neighborhoods, and cities), *geographical spaces* (e.g., regions and countries), *panoramic spaces*, and *map spaces* (including models) (cf. [128]). In this way, they cover existing as well as imagined places, where facts about the former can be mixed with imaginations of the latter [129]. This list includes the kind of places that are central to our study, especially cities.

To build a bridge between the notion of CMs and our analysis, we need to look more closely at their content and the principles by which they are created. Generally speaking, CMs are seen to cover at least two types of information (see [124] 1p. and [129] 314p.):

(1) Regarding *spatial cognition*, this concerns information about where entities are located in the environment of a person (location, distance, and direction in relation to her location or to reference points like landmarks)

(2) Regarding *environmental cognition*, this concerns information about the kind of these entities, their attributes, meanings, valuations, and attitudes that the person associates with them—individually, socially, or culturally mediated ([126], 224, 235)

Our study focuses on the second part of this distinction: it is related to the rhemes that are associated with places as framing themes (see Section 1). In any event, CMs are systematically characterized by distortions ([129], 315) concerning judgments about locations, distances, and directions as well as the formation of preferences which affect spatial or environmental cognition. One example is the *localization effect* [120] according to which people can discriminate nearby places better and have stronger preferences for them, see also [126]. This relates to errors in distance judgments depending on the perspective from which they are made: more differences are seen between closer areas than between more distant ones, so that shorter distances are exaggerated, while longer distances are underestimated [130], 133). Furthermore, spatial knowledge can be organized by reference to landmarks which "distort" places in their "neighborhood" so that buildings, for example, are judged to be closer to them than vice versa [130], 134). Tversky ([130], 135pp.] describes additional modes of distortion: to remember the position and orientation of objects, humans isolate them from their background and organize them by referring to a general frame of reference (rotation) or to other figures (alignment). While these examples primarily concern spatial cognition, the following bias focuses more on environmental cognition. This concerns the hierarchical organization of conceptual systems according to which places of the same category are supposed to be closer in distance than places of different categories, while the direction of a category (with a direction slot) determines the one of its members ([130], 132p). Last but not least, Golledge and Stimson [126] describe distortions of the representation of urban spaces. They observe that interactions influence the perception of a city in the sense that spatial information accumulates along the representations of the paths used to carry out these interactions. Likewise, structural properties of cities which are more salient than others are likely to become anchor points in CMs. In such maps, areas between used paths and anchor points may appear to be "folded" or "wrapped" so that preferred visited places are represented closer to each other. As a result, positional and relational errors can occur in perception (see ([126], 254) and ([131], 7).

To interpret our findings in the light of this research, we need to link the formation of CMs with linguistic processes. The idea that this formation is substantially influenced by human language processing, so that geographical information is nontrivially encoded in linguistic structure, goes back to the work of Louwerse and Benesh (cf. [26]) (see Section 1; see also Montello and Freundschuh ([132], 171)

for an earlier hint on "*obtain[ing] spatial knowledge through language*"). In this context, Golledge and Stimson ([126], 235) distinguish shared components of CMs from personalized ones by stating that "*The common elements facilitate communication with others about the characteristics of an environment; the idiosyncratic elements provide the basis of the personalized responses to such situations*." Our hypothesis is now that at the level of thematic abstraction as modeled here, the organization of platial rhemes shared by the members of a community is influenced by the general law of preferential order, which is most prominently instantiated by Zipf's first law [133]. Such an organization makes the anticipation of a place rather expectable among the members of a community so that communication about this place is facilitated as predicted by Golledge and Stimson [126].

This Zipfian organization allows for relating our findings to the well-known power-law-like degree distributions found in many natural, social, semiotic, or technical networks (see [109, 134] and especially [135] for overviews of this and related research) and also by example of many linguistic systems—especially on the text level [136–138]. Because of this commonality, one might assume that we just detected a well-known *text* or *network* characteristic. Characteristic for our findings, however, is that we developed a measurement procedure that detects a *text (corpus)-related semantic, thematic trend*—with the help of network theory: instead of counting directly observable arcs, for example, in ontological networks or co-occurrences in texts and instead of relying on monoplex networks [70, 93, 139–143], we generated and analyzed a range of different networks in relation to each other in order to determine the corresponding thematic trend by means of multiplex networks. This is not to say that we first discovered a Zipfian process in the organization of linguistic networks, but rather that we observe such a process in a very specific area, in which it has not been observed before and which requires an appropriate explanation as elaborated so far. Indeed, if thematic salience is skewed, and if skewed topic distributions derived from different corpora are similar not only topologically but also regarding the ranking of the majority of salient topics, such an observation requires explanation subject to the fact that the underlying text networks are constituted by different, distributed communities of authors. It is the answer to this question that the paper was about.

At this point, one might further object that we made a rather expectable observation in the sense that descriptions of cities, for example, are very likely related to rhemes like traffic, trade, culture, and history. However, this would mean underestimating our results: (i) the thematic distortions observed by us are extremely skewed, (ii) they seem to emerge rather earlier in the development of a wiki (this is not shown here but is the result of a pretest in which we looked at the life cycles of three different wikis; in future work, we will analyze the underlying time series of multiplex topic networks in detail), and (iii) they make both members of the same genre similar while allowing for distinguishing members of different genres. To phrase it as a question: *If the number of rhemes under which places are thematized is limited, why then should always a tiny subset of them dominate the discourse about a place and why then should the networking of these rhemes make discourses of the same genre identifiable?* From this point of view, we argue that we discovered an additional form of the distortion of CMs, which means that the underlying place is always conceptualized from the point of view of a few but extremely preferred rhemes. When organizing their distributed processes of coauthorship, communities of authors seem to strive to a kind of thematic unification that makes different wikis serving alike functions looking structurally similar—with respect to the preference order of themes and their networking. It seems that people participate in processes of collaborative writing with a tendency to organize their thematic contributions and references in such a way that they remain shareable [144] and communicable among members of the same community. Ensuring shareability means securing the continued existence of the underlying wiki, which could otherwise collapse because of too many personalized or individualized fragmentations. At this point, we can speculate that people unconsciously prefer such thematic contributions that make their social roles and participations expectable and acceptable, whereby this selection behavior produces the described similarity of thematic maps as components of CMs. In other words, the participants anticipate social roles and neglect their personal view of cities and regions, whose documentation would fragment the corresponding media thematically. Instead, they ignore the reproduction of their idiosyncratic, personalized views of places. To say it in terms of the distinction made by Golledge and Stimson [126] between shared and personalized components of CMs: participants overweight the former to the disadvantage of the latter to guarantee the shareability [144, 145] of CMs as a result of distributed cognition.

Note that in our study we did not simply map a frequency effect by our measurements: although we counted frequencies of topic assignments, they were determined by means of an inference process that went through a process of (machine) learning. To support such an interpretation, however, a deeper analysis with a larger corpus of wikis and related media providing different functions is required. This also requires experiments with other and above all much finer classification systems than the DDC to find out how much the use of the DDC has influenced our measurements. And it requires a deeper analysis of the social roles of authors in online media, their interactions, and the regulatory systems under which they interact. But this already concerns future work.

6. Conclusion

We developed a novel model of topic networks in order to investigate the networking of rhemes addressing the same places in underlying corpora of natural language texts. We developed our network model in a way that it enables thematic comparisons of previously unforeseen text corpora using an underlying reference corpus, offers a generic solution to the problem of topic labeling, is highly scalable and can therefore map even the smallest text snippets to topic

000	Computers, internet & systems	100	Philosophy	200	Religion	300	Social sciences, sociology & anthropology
010	Bibliographies	110	Metaphysics	210	Philosophy & theory of religion	310	Statistics
020	Library & information science	120	Epistemology	220	The Bible	320	Political science
030	Encyclopedias & books of facts	130	Astrology, parapsychology & the occult	230	Christianity & christian theology	330	Economics
040	Unassigned	140	Philosophical schools of thought	240	Christian practice & observance	340	Law
050	Magazines, journals & serials	150	Psychology	250	Christian pastoral practice & religious orders	350	Public administration & military science
060	Associations, organizations & museums	160	Logic	260	Church organization, social work & worship	360	Social problems & social services
070	Journalism, publishing & news media	170	Ethics	270	History of christianity	370	Education
080	Quotations	180	Ancient, medieval & Eastern philosophy	280	Christian denominations	380	Commerce, communications & transportation
090	Manuscripts & rare books	190	Modern western philosophy	290	Other religions	390	Customs, etiquette & folklore
400	Language	500	Science	600	Technology	700	Arts
410	Linguistics	510	Mathematics	610	Medicine	710	Landscaping & area planning
420	English & old English languages	520	Astronomy	620	Engineering	720	Architecture
430	German & related languages	530	Physics	630	Agriculture	730	Sculpture, ceramics & metalwork
440	French & related languages	540	Chemistry	640	Home & family management	740	Drawing & decorative arts
450	Italian, Romanian & related languages	550	Earth sciences & geology	650	Management & public relations	750	Painting
460	Spanish & Portuguese languages	560	Fossils & prehistoric life	660	Chemical engineering	760	Graphic arts
470	Latin & Italic languages	570	Biology & life sciences	670	Manufacturing	770	Photography
480	Classical & modern Greek languages	580	Plants (Botany)	680	Manufacturing specific products	780	Music
490	Other languages	590	Animals (Zoology)	690	Building & construction	790	Sports, games & entertainment
800	Literature, rhetoric & criticism	900	History				
810	American literature in English	910	Geography & travel				
820	English & old English literatures	920	Biography & genealogy				
830	German & related literatures	930	History of the ancient world (to ca. 499 A.D.)				
840	French & related literatures	940	History of Europe (ca. 500 A.D. -)				
850	Italian, Romanian & related literatures	950	History of Asia				
860	Spanish & Portuguese literatures	960	History of Africa				
870	Latin & Italic literatures	970	History of North America				
880	Classical & modern Greek literatures	980	History of South America				
890	Other literatures	990	History of other regions				

Figure 27: Color codes of the classes of the 2nd level of the DDC.

distributions, simultaneously takes rare topics into account, and is methodologically open and expandable. Moreover, our model allows for comparatively investigating the networking of thematic units from different angles. In this way, it is open and expandable as it allows for integrating different analytical perspectives into the study of the same semantic networks. We exemplified our model by means of corpora of special wikis and extracts from Wikipedia in order to investigate how textual information encodes geographical information on the aboutness level of texts. Our experiments show that the thematizations of different places on a certain level of abstraction are similar to each other in that they focus on a few themes in a highly distorted manner while networking them in similar ways. This happens regardless of whether the underlying media are generated by different communities and whether these communities address related or unrelated places in nearby or distant places. We interpreted our findings in the context of the notion of cognitive maps. To this end, we proposed to extend this notion in terms of thematic maps and argued that participants or interlocutors of online communication tend to organize their contributions in a way that makes them sharable. This means that the contributions are abstracted and depersonalized at the aboutness level in such a way that the social roles of these participants become expectable and acceptable, while their personal views of places are reduced whose documentation would fragment the corresponding media thematically. Ensuring shareability means securing the continued existence of the wiki, which could otherwise collapse in the face of too many personalized or individualized fragmentations. Future work concerns several tasks: we want to conduct deeper analyses based on larger corpora that manifest a greater variety of communication functions in order to shed more light on the genre sensitivity discovered in our study. Beyond the DDC, we strive for the use of finer structured, higher resolution classification systems in order to model the contents of texts much more precisely. Ideally, this should be carried out with the help of systems like the category system of Wikipedia or even Wikidata, both of which develop as open topic universes [146]. Last but not least, a deeper analysis of the social roles of authors in online media and their coauthorship is required to gain a deeper understanding of the processes of linguistic encoding of geographical information. This will be the task of future work.

Appendix

A. text2ddc

text2ddc is trained by means of corpora that are derived by integrating information from Wikidata, Wikipedia, and the *Integrated Authority File* (*Gemeinsame Normdatei*—GND) of the German National Library: we explore the links of Wikipedia articles to entries in Wikidata containing the property attribute https://www.wikidata.org/wiki/Property:P1036 that directly links to the DDC or to a GND page containing a DDC tag. An example is the article about the *Pythagorean theorem* (https://en.wikipedia.org/wiki/Pythagorean_theorem), which is linked to the GND

page 4176546-1 (https://d-nb.info/gnd/4176546-1) referring to the DDC tag 516 (*geometry*). Using such information, we obtain a corpus for a subset of 98 classes of the 2nd and for a subset of 641 classes of the 3rd DDC level. Since Wikipedia exists for many languages, such corpora can be created for each of them. For preprocessing the input data of text2ddc, we use TextImager [86] and fastSense [88] for disambiguating these data on the sense level. The resulting information is used to train a neural network for classifying any piece of text (down to the word level) into DDC classes (see https://textimager.hucompute.org/DDC/). To this end, text2ddc uses a very efficient classifier, that is, fastText [91], a bag-of-words model to train a neural network with a single hidden layer. To optimize fastText, we optimize the following hyperparameters: learning rate: 0; update rate: 150; minimal number of word occurrences: 5; number of epochs: 10,000. In this way, we increase the *F*-score to 87% for the 2nd level and to 78% for the 3rd level of the DDC.

B. Color Codes and 2nd Class Members of the DDC

Figure 27 shows the colors and labels of the classes of the 2nd level of the DDC.

Data Availability

Parts of the programs that underlie our work are available via GitHub (https://github.com/texttechnologylab/GeneticClassifierWorkbench).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Financial support by the Federal Ministry of Education and Research (BMBF) via the *Centre for the Digital Foundation of Research in the Humanities, Social, and Educational Sciences* CEDIFOR is gratefully acknowledged.

References

[1] A. Crooks, D. Pfoser, A. Jenkins et al., "Crowdsourcing urban form and function," *International Journal of Geographical Information Science*, vol. 29, no. 5, pp. 720–741, 2015.

[2] M. Crang and S. Graham, "SENTIENT CITIES ambient intelligence and the politics of urban space," *Information, Communication & Society*, vol. 10, no. 6, pp. 789–817, 2007.

[3] H. Chen, M. Vasardani, S. Winter, and M. Tomko, "A graph database model for knowledge extracted from place descriptions," *ISPRS International Journal of Geo-Information*, vol. 7, no. 6, p. 221, 2018.

[4] M. F. Goodchild, "Citizens as sensors: the world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, 2007.

[5] M. F. Goodchild and L. Li, "Assuring the quality of volunteered geographic information," *Spatial Statistics*, vol. 1, pp. 110–120, 2012.

[6] D. Hardy, J. Frew, M. F. Goodchild, and Goodchild, "Volunteered geographic information production as a spatial process," *International Journal of Geographical Information Science*, vol. 26, no. 7, pp. 1191–1212, 2012.

[7] D. Z. Sui, "The wikification of GIS and its consequences: or Angelina Jolie's new tattoo and the future of GIS," *Computers, Environment and Urban Systems*, vol. 32, no. 1, pp. 1–5, 2008.

[8] B. Jiang and J.-C. Thill, "Volunteered geographic information: towards the establishment of a new paradigm," *Computers, Environment and Urban Systems*, vol. 53, pp. 1–3, 2015.

[9] M. M. Salvini and S. I. Fabrikant, "Spatialization of user-generated content to uncover the multirelational world city network," *Environment and Planning B: Planning and Design*, vol. 43, no. 1, pp. 228–248, 2016.

[10] B. J. Hecht and D. Gergle, "On the "localness" of user-generated content," in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW '10*, pp. 229–232, ACM, New York, NY, USA, February 2010.

[11] M. Graham, B. Hogan, R. K. Straumann, and A. Medhat, "Uneven geographies of user-generated information: patterns of increasing informational poverty," *Annals of the Association of American Geographers*, vol. 104, no. 4, pp. 746–764, 2014.

[12] S. Hahmann and D. Burghardt, "How much information is geospatially referenced? Networks and cognition," *International Journal of Geographical Information Science*, vol. 27, no. 6, pp. 1171–1189, 2013.

[13] A. Jenkins, A. Croitoru, A. T. Crooks, and S. Anthony, "Crowdsourcing a collective sense of place," *PLoS One*, vol. 11, no. 4, Article ID e0152932, 2016.

[14] D. Clare, "Are places concepts? familarity and expertise effects in neighborhood cognition," in *Kathleen Stewart Hornsby, Christophe Claramunt, Michel Denis, and Gérard Ligozat, Spatial Information Theory*, pp. 36–50, Springer, Berlin, Heidelberg, 2009.

[15] A. Mehler, R. . Gleim, A. Lücking, T. Uslu, and C. Stegbauer, "On the self-similarity of Wikipedia talks: a combined discourse-analytical and quantitative approach," *Glottometrics*, vol. 40, pp. 1–45, 2018.

[16] M. M. Louwerse and R. A. Zwaan, "Language encodes geographical information," *Cognitive Science*, vol. 33, no. 1, pp. 51–73, 2009.

[17] D. L. Medin, R. L. Goldstone, and D. Gentner, "Respects for similarity," *Psychological Review*, vol. 100, no. 2, p. 254, 1993.

[18] K. Adamzik, *Textlinguistik: Grundlagen, Kontroversen, Perspektiven*, De Gruyter, Berlin, Germany, 2016.

[19] S. Yablo, *Aboutness*, Princeton University Press, Princeton, NJ, USA, 2014.

[20] K. Brinker, *Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden*, Erich Schmidt, Berlin, Germany, 1992.

[21] F. Daneš, "The paragraph—A central unit of the thematic and compositional build-up of texts," in *Organization in Discourse*, B. Wårvik, S.-K. Tanskanen, and R. Hiltunen, Eds., pp. 29–40, University of Turku, Turku, Finland, 1995.

[22] L. Hoffmann, "Thema, themenentfaltung, makrostruktur," in *Text- und Gesprächslinguistik/Linguistics of Text and Conversation—Ein internationales Handbuch zeitgenössischer Forschung*, K. Brinker, G. Antos, W. Heinemann, and

S. F. Sager, Eds., vol. 1, pp. 344–355, De Gruyter, Berlin, Germany, 2000.

[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[24] D. Clare, "Reading geography between the lines: extracting local place knowledge from text," in *Spatial Information Theory*, T. Tenbrink, John Stell, A. Galton, and Z. Wood, Eds., pp. 320–337, Springer International Publishing, Cham, Switzerland, 2013.

[25] B. B. Rieger, "Semiotic cognitive information processing: learning to understand discourse. A systemic model of meaning constitution," in *Adaptivity and Learning. An Interdisciplinary Debate*, R. Kühn, R. Menzel, W. Menzel, U. Ratsch, M. M. Richter, and I. O. Stamatescu, Eds., pp. 347–403, Springer, Berlin, Germany, 2003.

[26] M. M. Louwerse and N. Benesh, "Representing spatial structure through maps and language: lord of the rings encodes the spatial structure of middle earth," *Cognitive Science*, vol. 36, no. 8, pp. 1556–1569, 2012.

[27] W. R. Tobler, "A computer movie simulating urban Growth in the Detroit region," *Economic Geography*, vol. 46, pp. 234–240, 1970.

[28] H. J. Miller, "Tobler's first law and spatial analysis," *Annals of the Association of American Geographers*, vol. 94, no. 2, pp. 284–289, 2004.

[29] D. R. Montello, S. I. Fabrikant, M. Ruocco, and R. S. Middleton, "Testing the first law of cognitive geography on point-display spatializations," in *Spatial Information Theory. Foundations of Geographic Information Science*, W. Kuhn, M. F. Worboys, and S. Timpf, Eds., pp. 316–331, Springer, Berlin, Germany, 2003.

[30] B. Hecht and E. Moxley, "Terabytes of Tobler: evaluating the first law in a massive, domain-neutral representation of world knowledge," in *Spatial Information Theory*, K. S. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, Eds., pp. 88–105, Springer Berlin Heidelberg, Berlin, Germany, 2009.

[31] T. J.-J. Li, S. Sen, and B. Hecht, "Leveraging advances in natural language processing to better understand Tobler's first law of geography," in *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '14*, pp. 513–516, ACM, Fort Worth, TX, USA, November 2014.

[32] H. Yang, S. Chen, M. R. Lyu, and I. King, "Location-based topic evolution," in *Proceedings of the 1st International Workshop on Mobile Location-Based Service (MLBS'11)*, pp. 89–98, ACM, Beijing, China, September 2011.

[33] M. M. Louwerse, "Symbol interdependency in symbolic and embodied cognition," *Topics in Cognitive Science*, vol. 3, no. 2, pp. 273–302, 2011.

[34] B. Adams and G. McKenzie, "Inferring thematic places from spatially referenced natural language descriptions," in *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, D. Sui, S. Elwood, and M. Goodchild, Eds., pp. 201–221, Springer, Dordrecht, Netherlands, 2013.

[35] Q. Mei, C. Liu, H. Su, and C.X. Zhai, "A probabilistic approach to spatiotemporal theme pattern mining on weblogs," in *Proceedings of the 15th International Conference on World Wide Web, WWW '06*, pp. 533–542, ACM, Edinburgh, UK, May 2006.

[36] H. Qiang, R. Cai, C. Wang et al., "Equip tourists with knowledge mined from travelogues," in *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pp. 401–410, ACM, Raleigh, NC, USA, April 2010.

[37] A. R. Bahrehdar and R. S. Purves, "Description and characterization of place properties using topic modeling on georeferenced tags," *Geo-Spatial Information Science*, vol. 23, no. 3, pp. 173–184, 2018.

[38] Z. Yin, L. Cao, J. Han, C. Zhai, and T. Huang, "Geographical topic discovery and comparison," in *Proceedings of the 20th International Conference on World Wide Web (WWW'11)*, pp. 247–256, ACM, Hyderabad, India, March-April 2011.

[39] R. Gabriel and M. Max, "Louwerse. Grounding the ungrounded: estimating locations of unknown place names from linguistic associations and grounded representations," in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, CogSci 2014*, Quebec City, Canada, July 2014.

[40] M. Speriosu, T. Brown, T. Moon, J. Baldridge, and K. Erk, "Connecting language and geography with region-topic models," in *Proceedings of the Workshop on Computational Models of Spatial Language Interpretation (COSLI)*, vol. 46, Portland, OR, USA, August 2010.

[41] S. Gao, K. Janowicz, and H. Couclelis, "Extracting urban functional regions from points of interest and human activities on location-based social networks," *Transactions in GIS*, vol. 21, no. 3, pp. 446–467, 2017.

[42] G. Lansley and P. A. Longley, "The geography of Twitter topics in London," *Computers, Environment and Urban Systems*, vol. 58, pp. 85–96, 2016.

[43] S. Gao, K. Janowicz, D. R. Montello et al., "A data-synthesis-driven method for detecting and extracting vague cognitive regions," *International Journal of Geographical Information Science*, vol. 31, no. 6, pp. 1245–1271, 2017.

[44] D. R. Montello, "Regions in geography: process and content," *Foundations of Geographic Information Science*, pp. 173–189, 2003.

[45] E. Leopold, "Models of semantic spaces," in *Aspects of Automatic Text Analysis, Volume 209 of Studies in Fuzziness and Soft Computing*, A. Mehler and R. Köhler, Eds., pp. 117–137, Springer, Berlin, Germany, 2007.

[46] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211–240, 1997.

[47] C. Davies and T. Tenbrink, "Place as location categories: learning from language," in *Proceedings of Workshops and Posters at the 13th International Conference on Spatial Information Theory (COSIT 2017)*, P. Fogliaroni, A. Ballatore, and E. Clementini, Eds., pp. 217–225, Springer International Publishing, Cham,Switzerland, 2018.

[48] Y. Hu, X. Ye, and S.-L. Shaw, "Extracting and analyzing semantic relatedness between cities using news articles," *International Journal of Geographical Information Science*, vol. 31, no. 12, pp. 2427–2451, 2017.

[49] Y. Liu, F. Wang, C. Kang, Y. Gao, and Y. Lu, "Analyzing relatedness by toponym Co-occurrences on web pages," *Transactions in GIS*, vol. 18, no. 1, pp. 89–107, 2014.

[50] D. Ramage, D. Hall, N. Ramesh, D. Christopher, and Manning, "Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing EMNLP '09*, vol. 1, pp. 248–256, Association for Computational Linguistics, Singapore, August 2009.

[51] L. Björneborn, *Small-World link structures across an academic web space: a library and information science approach*, Ph.D. thesis, Royal School of Library and Information

Science, Department of Information Studies, Aalborg, Denmark, 2004.

[52] W. Luo, Y. Wang, X. Liu, and S. Gao, "Cities as spatial and social networks: towards a spatio-socio-semantic analysis framework," in *Cities as Spatial and Social Networks*, X. Ye and X. Liu, Eds., pp. 21–37, Springer International Publishing, Cham, Switzerland, 2019.

[53] P. Agarwal, "Operationalising 'sense of place' as a cognitive operator for semantics in place-based ontologies," in *Spatial Information Theory*, A. G. Cohn and D. M. Mark, Eds., pp. 96–114, Springer, Berlin, Heidelberg, 2005.

[54] S. Winter and C. Freksa, "Approaching the notion of place by contrast," *Journal of Spatial Information Science*, vol. 2012, no. 5, pp. 31–50, 2012.

[55] Y. Hu, "Geospatial semantics," 2017, https://arxiv.org/abs/1707.03550.

[56] M. Dehmer, A. Mehler, and Frank Emmert-Streib, "Graph-theoretical characterizations of generalized trees," in *Proceedings of the 2007 International Conference on Machine Learning: Models, Technologies & Applications (MLMTA'07)*, pp. 113–117, Las Vegas, NV, USA, June 2007.

[57] A. Mehler, "Structural similarities of complex networks: a computational model by example of wiki graphs," *Applied Artificial Intelligence*, vol. 22, no. 7-8, pp. 619–683, 2008.

[58] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[59] M. Steyvers, T. Griffiths, T. K. Landauer, D. S. McNamara, S. Dennis, and K. Walter, "Probabilistic topic models," in *Handbook of Latent Semantic Analysis*, pp. 427–448, Lawrence Erlbaum Associates, New York, NY, USA, 2007.

[60] G. Heinrich, *A generic approach to topic models and its application to virtual communities*, Ph.D thesis, University of Leipzig, Leipzig, Germany, 2012.

[61] B.-J. (Paul) Hsu and J. Glass, "Style & topic language model adaptation using HMM-LDA," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP' 06*, pp. 373–381, Association for Computational Linguistics, Sydney, Australia, July 2006.

[62] M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth, and M. Steyvers, "Learning author-topic models from text corpora," *ACM Transactions on Information Systems*, vol. 28, no. 1, pp. 1–38, 2010.

[63] A. Herzog, P. John, and S. J. Mikhaylov, "Transfer topic labeling with domain-specific knowledge base: an analysis of UK house of commons speeches 1935-2014," 2018, https://arxiv.org/abs/1806.00793.

[64] S. Boccaletti, G. Bianconi, R. Criado et al., "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014.

[65] M. Stella, N. M. Beckage, M. Brede, and M. De Domenico, "Multiplex model of mental lexicon reveals explosive learning in humans," *Scientific Reports*, vol. 8, no. 1, p. 2259, 2018.

[66] M. A. K. Halliday and R. Hasan, *Language, Context, and Text: Aspects of Language in a Socialsemiotic Perspective*, Oxford University Press, Oxford, UK, 1989.

[67] R. Clarke, "The persistence of systems in organisations," in *Signs of Work. Semiosis and Information Processing in Organisations*, B. Holmqvist, P. B. Andersen, H. Klein, and R. Posner, Eds., pp. 59–91, De Gruyter, Berlin, Germany, 1996.

[68] E. Ventola, *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters*, Pinter, London, UK, 1987.

[69] A. Mehler, "Generalized shortest paths trees: a novel graph class applied to semiotic networks," in *Analysis of Complex Networks: From Biology to Linguistics*, M. Dehmer and F. Emmert-Streib, Eds., pp. 175–220, Wiley-VCH, Weinheim, Germany, 2009.

[70] A. Mehler, "Social ontologies as generalized nearly acyclic directed graphs: a quantitative graph model of social ontologies by example of Wikipedia," in *Towards an Information Theory of Complex Networks: Statistical Methods and Applications*, M. Dehmer, F. Emmert-Streib, and A. Mehler, Eds., pp. 259–319, Birkhäuser, Basel, Switzerland, 2011.

[71] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.

[72] T. Uslu, A. Mehler, A. Niekler, and D. Baumartz, "Towards a DDC-based topic network model of wikipedia," in *Proceedings of 2nd International Workshop on Modeling, Analysis, and Management of Social Networks and their Applications (SOCNET 2018)*, February 2018.

[73] T. Uslu and A. Mehler, "PolyViz: a visualization system for a special kind of multipartite graphs," in *Proceedings of the IEEE VIS 2018 (IEEE VIS 2018)*, Berlin, Germany, October 2018.

[74] U. Brandes, P. Kenis, J. . Lerner, and D. van Raaij, "Network analysis of collaboration structure in Wikipedia," in *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*, pp. 731–740, ACM, Madrid, Spain, April 2009.

[75] M. E. J. Newman, "Who is the best connected scientist? A study of scientific coauthorship networks," in *Complex Networks*, E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, Eds., pp. 337–370, Springer, Berlin Germany, 2004.

[76] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[77] T. Mikolov, W.-tau Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in *Proceedings of the NAACL 2013*, pp. 746–751, Redmond, WA, USA, January 2013.

[78] S. Harispe, S. Ranwez, S. Janaqi, and J. Montmain, "Semantic similarity from natural language and ontology analysis," *Synthesis Lectures on Human Language Technologies*, vol. 8, no. 1, pp. 1–254, 2015.

[79] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proceedings of the National Academy of Sciences*, vol. 101, no. 1, pp. 5200–5205, 2004.

[80] A. Komninos and S. Manandhar, "Dependency based embeddings for sentence classification tasks," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1490–1500, San Diego, CA, USA, June 2016.

[81] O. Levy and Y. Goldberg, "Dependency-based word embeddings," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 2, pp. 302–308, Baltimore, MD, USA, June 2014.

[82] L. Wang, C. Dyer, A. Black, and I. Trancoso, "Two/too simple adaptations of word2vec for syntax problems," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Denver, CO, USA, June 2015.

[83] A. Budanitsky and G. Hirst, "Evaluating WordNet-based measures of lexical semantic relatedness," *Computational Linguistics*, vol. 32, no. 1, pp. 13–47, 2006.

[84] W. Hemati, T. Uslu, and A. Mehler, "TextImager: a distributed UIMA-based system for NLP," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pp. 59–63, Osaka, Japan, December 2016.

[85] B. Bohnet, J. Nivre, I. Boguslavsky, R. Farkas, F. Ginter, and J. Hajič, "Joint morphological and syntactic analysis for richly inflected languages," *Transactions of the Association for Computational Linguistics*, vol. 1, pp. 415–428, 2013.

[86] T. Uslu, A. Mehler, D. Baumartz, H. Alexander, and W. Hemati, "Fastsense: an efficient word sense disambiguation classifier," in *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC 2018)*, Miyazaki, Japan, May 2018.

[87] S. Eger, R. . Gleim, and A. Mehler, "Lemmatization and morphological tagging in German and Latin: a comparison and a survey of the state-of-the-art," in *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, Portorož, Slovenia, May 2016.

[88] D. Baumartz, T. Uslu, and A. Mehler, "LTV: labeled topic vector," in *Proceedings of the COLING 2018, the 26th International Conference on Computational Linguistics: System Demonstrations*, The COLING 2018 Organizing Committee, Santa Fe, New Mexico, USA, August 2018.

[89] J. Armand, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," 2016, https://arxiv.org/abs/1607.01759.

[90] M. Dehmer, "Information processing in complex networks: graph entropy and information functionals," *Applied Mathematics and Computation*, vol. 201, no. 1-2, pp. 82–94, 2008.

[91] P. Baldi, P. Frasconi, and P. Smyth, *Modeling the Internet and the Web*, Wiley, Chichester, UK, 2003.

[92] A. Mehler, R. . Gleim, W. Hemati, and T. Uslu, "Skalenfreie online soziale Lexika am Beispiel von Wiktionary," in *Proceedings of the 53rd Annual Conference of the Institut für Deutsche Sprache (IDS)*, S. Engelberg, H. Lobin, K. Steyer, and S. Wolfer, Eds., pp. 269–291, De Gruyter, Mannheim, Germany, March 2017.

[93] O. Abramov and A. Mehler, "Automatic language classification by means of syntactic dependency networks," *Journal of Quantitative Linguistics*, vol. 18, no. 4, pp. 291–336, 2011.

[94] L. Geng, M. Semerci, B. . Yener, J. Mohammed, and Zaki, "Graph classification via topological and label attributes," in *Proceedings of the 9th International Workshop on Mining and Learning with Graphs (MLG)*, San Diego, CA, USA, July 2011.

[95] T. Li, D. Han, Y. Shi, and M. Dehmer, "A comparative analysis of new graph distance measures and graph edit distance," vol. 403-404, , pp. 15–21, Information Sciences, 2017.

[96] M. Owen and W. Richards, "Graph comparison using fine structure analysis," in *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SOCIALCOM '10*, pp. 193–200, IEEE Computer Society, Minneapolis, Minnesota, USA, August 2010.

[97] F. Emmert-Streib, M. Dehmer, and Y. Shi, "Fifty years of graph matching, network alignment and network comparison," *Information Sciences*, vol. 346-347, pp. 180–197, 2016.

[98] D. Koutra, A. Parikh, A. Ramdas, and J. Xiang, "Algorithms for graph similarity and subgraph matching," 2011, https://www.cs.cmu.edu/jingx/docs/DBreport.pdf.

[99] D. Koutra, N. Shah, and T. Joshua, "Vogelstein, brian Gallagher, and Christos Faloutsos. DeltaCon: principled massive-graph similarity function with attribution," *ACM Trans. Knowl. Discov. Data*, vol. 10, no. 3, pp. 1–43, 2016.

[100] H. Bunke, P. J. Dickinson, M. Kraetzl, and W. D. Wallis, *A Graph-Theoretic Approach to Enterprise Network Dynamics (Progress in Computer Science and Applied Logic (PCS))*, Birkhäuser, Boston, MA, USA, 2006.

[101] R. Ibragimov, M. Malek, J. Guo, and B. Jan, "GEDEVO: an evolutionary graph edit distance algorithm for Biological network alignment," in *Volume 34 of OpenAccess Series in Informatics (OASIcs)*, T. Beißbarth, M. Kollmar, A. Leha et al., Eds., pp. 68–79, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2013.

[102] W. D. Wallis, P. Shoubridge, M. Kraetz, and D. Ray, "Graph distances using graph union," *Pattern Recognition Letters*, vol. 22, no. 6-7, pp. 701–704, 2001.

[103] P. Papadimitriou, D. Ali, and H. Garcia-Molina, "Web graph similarity for anomaly detection," in *Proceedings of the 17th International Conference on World Wide Web, WWW '08*, pp. 1167-1168, ACM, Beijing, China, April 2008.

[104] S. Adam, H. Bunke, M. Last, and K. Abraham, *Graph-Theoretic Techniques for Web Content Mining*, World Scientific, Singapore, 2005.

[105] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[106] P. Blanchard and D. Volchenkov, *Mathematical Analysis of Urban Spatial Networks*, Springer, Berlin, Germany, 2009.

[107] M. Berlingerio, D. Koutra, T. Eliassi-Rad, and C. Faloutsos, "Network similarity via multiple social theories," in *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1439-1440, IEEE, Niagara, Ontario, Canada, August 2013.

[108] S. Soundarajan, T. Eliassi-Rad, and B. Gallagher, "A guide to selecting a network similarity method," in *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 1037–1045, SIAM, Philadelphia, PA, USA, April 2014.

[109] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[110] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[111] B. Bollobás and O. M. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of Graphs and Networks. From the Genome to the Internet*, S. Bornholdt and H. G. Schuster, Eds., pp. 1–34, Wiley-VCH, Weinheim, Germany, 2003.

[112] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.

[113] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, 2005.

[114] G. Kalna, J. Desmond, and Higham, "Clustering coefficients for weighted networks," in *Proceedings of the Symposium on network analysis in natural sciences and engineering*, vol. 45, September 2006.

[115] R. Gleim, A. Mehler, Y. Sung, and S. WikiDragon, "A Java framework for diachronic content and network analysis of mediawikis," in *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference*, Miyazaki, Japan, May 2018.

[116] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[117] N. Ramli and D. Mohamad, "On the Jaccard index similarity measure in ranking fuzzy numbers," *Matematika*, vol. 25, pp. 157–165, 2009.

[118] E. C. Tolman, "Cognitive maps in rats and men," *Psychological Review*, vol. 55, no. 4, pp. 189–208, 1948.

[119] K. Lynch, *The Image of the City*, Vol. 11, MIT press, Cambridge, MA, USA, 1960.

[120] P. Gould and R. White, *Mental Maps*, Routledge, Abingdon, UK, 1986.

[121] J.J. Gieseking, "Where we go from here: the mental sketch mapping method and its analytic components," *Qualitative Inquiry*, vol. 19, no. 9, pp. 712–724, 2013.

[122] C. Helferich, "Mental maps und narrative raumkarten," in *Methoden der Kulturanthropologie*, K. Oehme-Jüngling and W. L. Christine Bischoff, Eds., pp. 241–256, Haupt Verlag, Bern, Switzerland, 2014.

[123] J. Portugali, "The construction of cognitive maps: an introduction," in *The Construction of Cognitive Maps*, pp. 1–7, Springer, Berlin, Germany, 1996.

[124] R. M. Kitchin, "Cognitive maps: what are they and why study them?," *Journal of Environmental Psychology*, vol. 14, no. 1, pp. 1–19, 1994.

[125] R. M. Downs and D. Stea, *Kognitive Karten. Die Welt in Unseren Köpfen*, Harper & Row, New York, NY, USA, 1982.

[126] R. G. Golledge and R. J. Stimson, *Spatial Behavior: A Geographic Perspective*, Guilford Press, New York, NY, USA, 1996.

[127] R. Kitchin and M. Blades, *The cognition of geographic space*, Vol. 4, Ib Tauris, London, UK, 2002.

[128] S. M. Freundschuh and M. J. Egenhofer, "Human conceptions of spaces: implications for GIS," *Transactions in GIS*, vol. 2, no. 4, pp. 361–375, 1997.

[129] R. M. Downs and David Stea, "Cognitive maps and spatial behaviour: process and product," in *The Map Reader: Theories of Mapping Practice and Cartographic Representation*, D. Martin, R. Kitchin, and C. Perkins, Eds., pp. 312–317, John Wiley & Sons, Hoboken, NJ, USA, 2011.

[130] B. Tversky, "Distortions in cognitive maps," *Geoforum*, vol. 23, no. 2, pp. 131–138, 1992.

[131] R. G. Golledge and T. GÃ¤rling, *Spatial Behavior in Transportation Modeling and Planning*, University of California Transportation Center, Berkeley, CA, USA, 2001.

[132] D. R. Montello and S. M. Freundschuh, "Sources of spatial knowledge and their implications for GIS: an introduction," *Geographical Systems*, vol. 2, no. 1, pp. 169–176, 1995.

[133] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*, Hafner Publishing Company, New York, NY, USA, 1972.

[134] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford UK, 2010.

[135] M. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.

[136] S. Naranan and V. K. Balasubrahmanyan, "Models for power law relations in linguistics and information science," *Journal of Quantitative Linguistics*, vol. 5, no. 1-2, pp. 35–61, 1998.

[137] A. Rapoport, "Zipf's law re-visited," in *Studies on Zipf's Law*, H. Guiter and M. V. Arapov, Eds., pp. 1–28, Brockmeyer, Halle, Germany, 1982.

[138] J. Tuldava, *Methods in Quantitative Linguistics*, Wissenschaftlicher Verlag, Trier, Germany, 1995.

[139] D. Raphael Amancio, O. N. Oliveira Jr., and L. F. Costa, "Identification of literary movements using complex networks to represent texts," *New Journal of Physics*, vol. 14, p. 43029, 2012.

[140] A. Baronchelli, R. Ferrer-i-Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen, "Networks in cognitive science," *Trends in Cognitive Sciences*, vol. 17, no. 7, pp. 348–360, 2013.

[141] C. Cattuto, A. Barrat, A. Baldassarri, G. Schehr, and V. Loreto, "Collective dynamics of social annotation," *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp. 10511–10515, 2009.

[142] R. F. i Cancho, A. Mehler, O. Pustylnikov, and A. Díaz-Guilera, "Correlations in the organization of large-scale syntactic dependency networks," in *Proceedings of the TextGraphs-2 at NAACL-HLT'07*, Rochester, New York, April 2007.

[143] R. F. i Cancho and R. V. Solé, "The small-world of human language," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 268, no. 1482, pp. 2261–2265, 2001.

[144] J. Freyd, "Shareability: the social psychology of epistemology," *Cognitive Science*, vol. 7, no. 3, pp. 191–210, 1983.

[145] J. J. Freyd, "What is shareability?," 2005, http://dynamic.uoregon.edu/jjf/defineshareability.html.

[146] A. Mehler and U. Waltinger, "Enhancing document modeling by means of open topic models," *Library Hi Tech*, vol. 27, no. 4, 2009.

Hindawi
Complexity
Volume 2019, Article ID 7690869, 17 pages
https://doi.org/10.1155/2019/7690869

Research Article

Network Growth Modeling to Capture Individual Lexical Learning

Nicole M. Beckage[1] **and Eliana Colunga**[2]

[1]*University of Wisconsin, Madison, USA*
[2]*University of Colorado, Boulder, USA*

Correspondence should be addressed to Nicole M. Beckage; nicolebeckage@gmail.com

Received 8 September 2018; Revised 22 February 2019; Accepted 25 July 2019; Published 31 October 2019

Academic Editor: Vincent Labatut

Network models of language provide a systematic way of linking cognitive processes to the structure and connectivity of language. Using network growth models to capture learning, we focus on the study of the emergence of complexity in early language learners. Specifically, we capture the emergent structure of young toddler's vocabularies through network growth models assuming underlying knowledge representations of semantic and phonological networks. In construction and analyses of these network growth models, we explore whether phonological or semantic relationships between words play a larger role in predicting network growth as these young learners add new words to their lexicon. We also examine how the importance of these semantic and phonological representations changes during the course of development. We propose a novel and significant theoretical framework for network growth models of acquisition and test the ability of these models to predict what words a specific child is likely to learn approximately one month in the future. We find that which acquisition model best fits is influenced by the underlying network representation, the assumed process of growth, and the network centrality measure used to relate the cognitive underpinnings of acquisition to network growth. The joint importance of representation, process, and the contribution of individual words to the predictive accuracy of the network model highlights the complex and multifaceted nature of early acquisition, provides new tools, and suggests experimental hypotheses for studying lexical acquisition.

1. Introduction

Children do not learn words in isolation. Instead children must learn the meanings and relationships of words within a communicative context, and in the context of other words. These same relationships, which make learning initial words challenging, likely offer scaffolding and support that help children make sense of the world around them. The connections and relationships between words likely aid future learning of new words. How the structure of language develops through the course of acquisition, and how this structure facilitates future language learning, is critical to understanding the acquisition process of early learners.

Here we set forth to build a predictive network growth model of the words a child is likely to learn next based on the emerging structure of the child's current vocabulary. We represent the child's current productive vocabulary as a network, with the words the child produces as nodes in the graph, and edges connecting words based on either semantic or phonological similarity. Our growth models assume that words enter the graph (become part of the child's productive vocabulary) based on either the child's current vocabulary knowledge or the structure of the global language environment as captured by the full network structure. With these assumptions, we have a systematic way of linking possible learning mechanisms and processes to the structure and connectivity of a child's lexical network.

Although it is unlikely that language is represented in the mind as a network, it is probable that the structure of a child's current vocabulary, or the structure of the language learning environment, influences language learning and lexical acquisition. Network representations provide a method for abstracting the complexity of language, allowing researchers to study the emergence of linguistic structure [1–4]. Here we make the simplifying assumption that language can be represented as a network and that we can learn

about the underlying cognitive process of acquisition by studying the change and growth of these language representations.

There is a rich amount of work considering both semantic and phonological networks of the adult lexicon (for a review, see [4–6]). Here we instead focus our review of previous work only on network analysis as related to the growing lexicon of early language learners. Structurally, early language networks show evidence of small-world structure, characterized by short average path lengths and high local clustering, even for very small graphs [1, 3, 7]. Interestingly, the small-world structure of these early lexical graphs has been shown to be correlated with language learning skills. In particular, 2-year-olds classified as late talkers, or those who have relatively small vocabularies compared to their peers, have been found to have lexical networks with less small-world structure than we would expect based on random network models [3]. This and other results suggest that the small-world structure of early language graphs is able to highlight or accentuate relevant features that may be important and facilitatory to future language learning, and thus small-world structure as observed in the growing lexical graphs may relate to the child's language learning ability.

Previous work has also linked topological features of language networks to the process of acquisition. Steyvers and Tenenbaum [1] proposed that language is learned by a process similar to preferential attachment, which we call *Preferential Growth* to remind the reader that the assumed mechanisms and process of growth vary from the formalization of preferential attachment [8], with highly connected nodes being learned earliest and those highly connected nodes facilitating the learning of new nodes based on their connectivity to these early learned highly connected nodes. The model, which simulates aspects of semantic differentiation, suggests that words are more likely to be learned if they connect to already known, well-connected words in the child's current lexical network [1]. Under semantic differentiation, new words or concepts are learned in relation to already known words [9–11]. In a network growth framework, a learned word attaches to highly connected nodes in the current vocabulary graph and then forms edges with the neighbors of the attachment node. The resulting network maintains the scale-free structure found in some semantic networks and also maintains high local clustering found in all language networks. This model suggests a mechanism for the observed empirical result of a strong correlation between the connectivity of words in semantic graphs and the reported age of acquisition of words [1].

Hills and colleagues [2, 7] suggest, instead, that language learning is driven by contextual diversity, or, in network terms, by the connectivity of unknown words in the language environment or the language of adults. Encoding adult language as a graph, connectivity of words may be related to structure in the environment (e.g., ball and catch), close proximity in spoken language (e.g., cat and dog), or even close proximity in physical space (e.g., chair and table). The connectivity of individual words in the full language graph may approximate the number of contexts and meanings of an individual word. Under a contextual diversity hypothesis, a word is more likely to be learned if it appears in multiple contexts since, with multiple exposures, the ambiguity of meaning will decrease [12, 13]. Experimental work has shown many ways in which multiple contexts and exposures can increase the likelihood of learning, as shown in cross-situational learning tasks [13], or via attentional mechanisms [14]. The model which operationalizes contextual diversity in network growth models is known as *Preferential Acquisition* and was proposed and validated in work by Hills and colleagues [2, 7, 15].

Here we propose a network modeling framework to predict the *individual words* a child is likely to learn next. While previous models have focused on normative acquisition [1, 2], we apply these growth models to the language trajectories of *individual children*. We additionally formalize a mathematical relationship between network analysis models of growth and the process of language acquisition in young children. The use of network models to explain acquisition requires (1) a clear definition of edges or similarity between words, (2) a systematic influence of the network structure on future vocabulary acquisition, and (3) a way of relating network structure to the acquisition of individual words. In the next section we formalize our framework for modeling individual acquisition before evaluating performance of the proposed models in the context of our framework. We find that each assumption (the graph, growth process, and conversion of network models to measures of to probability of learning a specific word) plays an important role, impacting our ability to accurately capture the language acquisition of individual children. We argue that this framework offers novel insight into the acquisition processes involved in early word learning. Most promising is that our models outperform baseline especially when predicting the vocabulary development trajectories of children who are learning language slower than their peers. We find a strong relationship between this improvement in performance and assumptions about growth mechanisms underlying our model. In future work, we aim to explore whether the predictive accuracy of these network models can motivate intervention and/or empirical investigations, providing novel insight into the processes that underlie language acquisition in young children.

2. Network Growth Modeling Framework

When explaining vocabulary acquisition with network growth models, we should consider all the aspects of the network structure that might influence predictability and interpretability of the acquisition process. Toward this goal, we propose the following three levels of analysis to frame and understand our models and their ability to capture individual learning trajectories:

(1) Macro level: the definition of a graph in terms of nodes and edges. Edges are based on measures of relatedness that may be thresholded by a criterion specifying when an edge exists between two nodes.

(2) Mezzo level: measure of (changing) structure and influence of structural properties of a chosen graph on the topic of study.

(3) Micro level: the interaction of nodes with other nodes. Here we investigated different centrality measures as a proxy for node importance.

In the case of child language acquisition, the strongest form of the hypothesis assumes that if we have (1) a meaningful definition of relationships between words, (2) a growth process that correctly approximates learning, and (3) a cognitively relevant measure of word importance, then we can accurately model the acquisition process. We evaluate our assumptions on the model's ability to predict the specific words an *individual child* is likely to learn next given the child's current language network.

Beyond the goal of accurate prediction, we can also investigate performance of our models at each of these levels of analysis. For example, at the macro level, we can ask, given the acquisition trajectories of individual children, if semantic or phonological features are more predictive in capturing the acquisition trajectory of new words. The difference between phonological similarity and semantic feature similarity has previously been considered, assuming that words are added based on a "rich get richer" or preferential growth model [16], but under this framework, we can extend these results by exploring the interaction of phonology and semantic features with various growth processes, capturing performance of the mezzo level. We also examine cognitive processes at the micro level by asking such questions as whether words that bridge the network (have high betweenness) or words that have many connections (high degree) are more likely to be learned earlier.

By considering each of these levels of analysis, we can better understand the interactions among these three levels and their effect on explaining the acquisition trends of young children. We evaluate the role of each of these levels in predictive modeling and use these results to inform our understanding of the processes that influence early acquisition. These network models are not only predictive models but may also suggest mechanisms and attentional influences that alter individual language trajectories. To preview our results, we find that all levels of analysis matter, but that there is a high amount of variability in ways in which children learn, suggesting that future research must focus on understanding the relationship between the network framework and individual cognitive and developmental differences of young children.

3. Methods

We define the probability of a word i being learned by child x as

$$\Pr(i \mid x) = \frac{1}{1 + \exp -\left(\beta_0 + \beta_1 \delta_{\mathrm{cdi}_{(i,x)}} + \beta_2 \delta_{(i,x)}\right)}, \quad (1)$$

which is a logistic transformation that includes an intercept term (β_0) and includes what we call the baseline word learnability for word i given child x ($\delta_{\mathrm{cdi}_{(i,x)}}$) and a model based measure or *growth value* for word i ($\delta_{(i,x)}$), conditioned on the vocabulary of child x. The model-based growth value is calculated based on network growth assumptions explained below. βs are free parameters that are learned in the training and validation portion of our models through standard logistic regression methods. In all models, we include $\delta_{\mathrm{cdi}_{(i,x)}}$ which is the baserate probability that word i is learned according to the CDI norms, renormalized for the known vocabulary of the child x and constrained by the underlying network representation used to compute the growth value. If child x already produces a word, we do not compute the probability of learning that word.

Under our network growth framework used to predict the probability of word i being learned by child x, the growth value $\delta_{(i,x)}$ is defined based on an assumed growth process and an assumed centrality (defined for three models in equations (3) through (5) which is naturally dependent on the underlying network representation). The centrality measure can be either local (e.g., degree as this considers only immediate neighbors) or global (e.g., betweenness as this weighs the contribution of all nodes in the graph). The model-dependent *growth value* $\delta_{(i,x)}$, which is minimally zero, is mapped to a probability through a logistic transformation characterized by scaling parameters β. Note that the above model allows for simultaneously predicting the set of words learned by an individual child, not only the single most likely word. Modeling learning of the single most likely word has been examined in a similar paradigm by Hills and colleagues using a mathematical formalism very similar to the model we propose here [2, 7]. βs are optimized across training data for each network representation, growth process, and centrality we compared and evaluated on unseen children and their vocabulary growth trajectory. $\delta_{\mathrm{cdi}_{(i,x)}}$ is computed the same way for each model and varies for individual children only based on renormalization of the CDI norms for that specific child x, as well as based on the words in the network representation used to compute $\delta_{(i,x)}$. We discuss $\delta_{\mathrm{cdi}_{(i,x)}}$ and define the growth values for each model in more detail below.

In our modeling framework, we focus on a subset of early learned words that is widely studied in the developmental literature. These words are part of a language development checklist, the MacArthur-Bates Communicative Development Inventory, which contains words that at least 50% of children know by the time they turn 2 years (e.g., CDIs [17]). This checklist is also normed such that we know, on average, what percent of children at a specific age produce each word. Although we further discuss the words on this list—and research using these words—below, we note that using a fixed set of words poses a unique inference challenge for our modeling approach. We need to simplify the full language graph to the words on the CDI. But some of the network representations do not contain information on the specific CDI words (e.g., "mommy" is on the CDI, not "mother," but "mother" is in more network representations than "mommy"). Given that network growth models cannot generalize to words outside of the network representation, we use the baseline age-specific CDI norms to compute

$\delta_{\text{cdi}}$and use the logistic model from the CDI norms as predictions of words outside the network. Thus, we have a basic ensemble model—if the network model does not predict the learning of a word in the CDI norms, we can use a logistic regression using the renormalized normative age of acquisition rate (δ_{cdi}) as the probability of learning that specific word. While we mention that fact here, for most of the analysis in this paper we focus on comparing the network representation and CDI baseline models only on those words that are in the specific network representation we are examining.

In our characterization of network growth models, we explore three different network representations. Specifically, we focus on semantic and phonological graphs. Semantic and phonological information are both known to affect the course of language development, but here we instead consider these representations as part of the generative mechanism. We evaluate the role of a specific network representation based on its ability to accurately predict the words learned by an individual child. We define $\delta_{(i,x)}$ to be a function of graph centrality, which is inherently dependent on the underlying network representation and the vocabulary of child x. We use model performance to explore whether semantic or phonological information is more useful in predicting individual acquisition trajectories. The network representations we explore are as follows:

Count of Shared Features from the McRae Feature Norms. The feature norms are based on open-ended features listed by participants for specific items [18]. A weighted edge exists indicating the proportion of shared features between two items. Construction mimics the network construction of Hills et al. [2, 7].

Nelson Free Association Norms [19]. An edge represents the proportion of individuals who, when given a cue word *a,* responded with word *b*. For example "cat" is a frequent response to the cue "dog"; in the network an edge exists from dog to cat with a weight proportional to response frequency.

Phonological Levenshtein Distance (Phonology) [20]. The inverse of the number of substitutions, insertions, or deletions required to transform the phonological form of one word to another word. For example, "dog" and "log" have an edit distance of 1 (by substitution).

3.1. Growth Modeling Equations. We consider three different network growth models originally proposed by Hills and colleagues [2] updated for our modeling framework. We formalize the growth models to provide mathematical equations and understanding as to how these models are distinct in the underlying growth mechanisms and in their ability to account for future lexical acquisition of individual children (see Figure 1 for a visual representation of predictions resulting from these models). For the mathematical formalization, each network graph can be represented as a an adjacency matrix N of size $|V|$, where V includes all unknown and known words of the child that are part of the CDI vocabulary checklist and network representation. $N_{i,j} \in [0, 1]$ indicates the existence and weight of an edge from i to j. We assume network representations are converted to binary such that $A_{i,j} \in \{0, 1\}$ (see below for motivation and more information). The threshold θ, for $A_{i,j} = N_{i,j} > \theta$, is learned specifically for each model we consider such that the growth model is maximally predictive on the training data. We also define the binary network induced by the set of words a child can produce at a particular point in time as K. c_j^M is the "importance" of node j, given network M. We presume edges of a network are defined by the chosen network definition. Edges form the basis for the calculation of a word's importance. We additionally assume that if a word is learned, the word and all respective edges to known words are added to the child's lexical graph. We do not consider words to be learned sequentially but instead predict learning of the joint set of words. This nonsequential assumption of our model is important as some growth processes assume that words are learned conditioned on the current vocabulary of the child, and thus the model estimates will differ based on the assumption of batch versus sequential learning. We choose batch learning as the data used to fit the model are themselves a set of words the child learns, and thus we avoid inference over the order in which these words are learned. We discuss this in more detail below.

All models make predictions only for unknown words. We define each network growth model as the function that maps c_i^M, the measure of the importance of word i, to δ_i, the probability of learning word i, for each unknown word i. For some models, M is based on the specific words known by child x, in which case we add an index x to δ_i and to M to indicate the role of that particular child's language knowledge. This δ value, or node "importance," simply provides a means to quantify the model's estimate of the utility that an individual (unknown) word has if learned. This is likely different for each specific child and is likely influenced by the words the child knows. In our work here, we consider different types of centralities and their relation to the lexical graph to operationalize the importance of individual words. Other types of word-specific measures, such as frequency or word length, could also be used.

Normative Age of Acquisition refers to our baseline estimate of learning approximated by the child's age, known vocabulary, and normative CDI reports. The normative CDI reports include, for each specific word, the percentage of children at a given age (rounded to the nearest month) that were reported to produce that word. We can then compute a $\delta_{\text{cdi}_{(i,x)}}$ value for each unknown word. This is computed by considering the child's age and linearly interpolating the normative CDI reports to calculate the percent of children at that given age who would produce the word.

Defining $B_{(i,m)}$ as the linearly interpolated proportion of children who know a particular word i when the child is m months old, we can compute the baserate probability of child x learning all unknown words on the CDI ($\delta_{\text{cdi}_{(i,x)}}$) for each word i

$$\delta_{\text{cdi}_{(i,x)}} = \frac{B_{(i,m_x)}}{\sum_j \delta_{\text{cdi}_{(j,x)}}}, \tag{2}$$

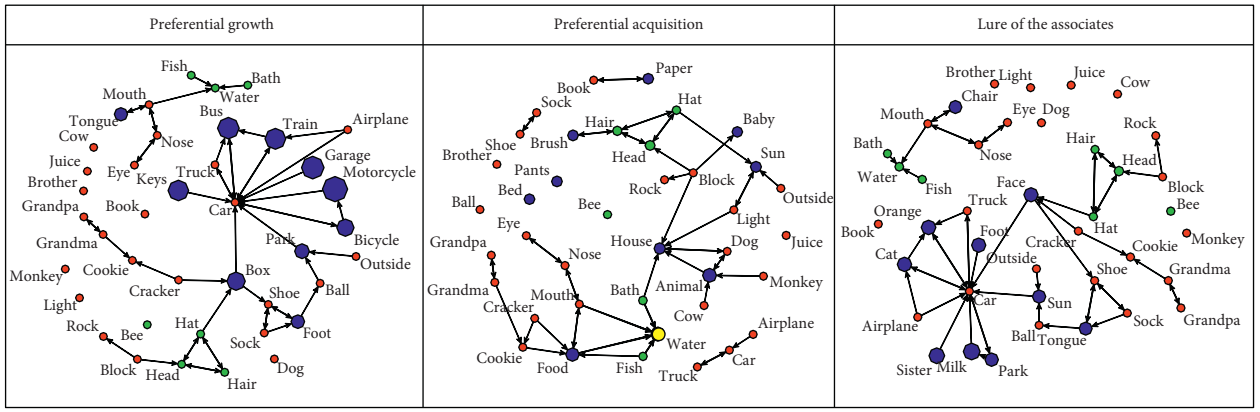


FIGURE 1: The three growth models with vocabulary predictions given a specific child's vocabulary. Red words are words already known by the child, green words are words learned by the child, blue words are the 10 most likely words to be learned according to model predictions, and the yellow word (water) is a word that was correctly predicted as learned by a specific model. The size of the blue dots indicates the certainty the model gives to learning that specific word.

where m_x is the age in months of child x at the point of prediction.

In all of our network growth models, we include this δ_{cdi} value. The performance of a model that only includes the δ_{cdi} value in equation (1) is then our baseline model. This baseline prediction can also used for all words not included in the network representation to allow comparison across network representations that include different words.

Preferential Growth assumes that words are more likely to be learned if they connect to nodes that are themselves well connected in the graph. A word is added to the graph, or learned by the child, in proportion to the total importance of each *known* word it attaches to. For example, looking at Figure 1, the word "car" is connected in the child's known vocabulary K_x, and thus this model assumes that words related to "car" are most likely to be learned next. Under this definition, the preferential growth model can be defined as

$$\delta_{i,x} \propto \sum_{j \in K_x} A_{i,j} c_j^{K_x}. \tag{3}$$

In this equation, we sum up the centrality $c_j^{K_x}$ of known words only if there is a connection between the unknown word i and j. We do this by using $A_{i,j}$ as an indicator to ensure that word i connects to word j. If word i is connected to word j, we consider word j as contributing some increase in the probability of learning word i. We define this increase in likelihood of learning as $c_j^{K_x}$ (the centrality measures of word j as computed from the network of words known by the child (K_x)). We do this for all known words. This sum is then considered proportional to the growth value of word i. For example, if we assume centrality is in-degree, this results in the sum of the in-degrees of the words (j) that i would connect to if learned. Note that the network of known words K is conditioned on the particular child's vocabulary at the time of prediction and thus includes the subscript for child id x.

This type of preferential growth mechanism would suggest that words are learned if they connect to highly important or central already known words. One possible cognitive mechanism that could drive this model is semantic or phonological differentiation, in which words are learned if they are similar to already known concepts (semantics) or sounds (phonology) [1]. One key feature of this model is that predictions of word learning is driven only by the (connectivity of) words that the child currently produces (K_x), with no influence of the global language environment.

Preferential Acquisition instead implies that words are learned based on their connectivity in the larger language environment, summarized in a network context as the "*full language graph*." We consider this full language graph to be the graph constructed presuming the child knows all the words in our vocabulary assessment. Mathematically, the growth value of each word under preferential acquisition is defined as follows:

$$\delta_i \propto c_i^A, \tag{4}$$

where A is the full binary network. This model relies on the idea that the more important a word is in the global environment, the earlier it is learned. Additionally, this model assumes that the full language graph approximates the *language environment* and *linguistic context* that is important to child language learning. What makes a word central varies on the specific network definition and centrality measure used to approximate the importance of individual words. For example, if we use in-degree centrality, this model assumes that words that have the most neighbors in the full language graph are those that are most likely to be learned next. Such a growth process might indicate that children are learning words that are contextually diverse (e.g., appear in a variety of context) and those words that are most likely to be learned earliest. This could be due to the fact that children can more easily learn word-object mappings when the object appears in many different environments [21], a fact that is naturally captured in the degree of a node in networks like free association graphs [2, 15]. Note that this model is not influenced by the child's vocabulary graph, and thus there is no need to have an index (x) for the child. Individual-level differences simply emerge from the fact that children know different words, and thus normalization will result in different probability distributions for individual children.

Lure of the Associates bridges the gap between a model based only on the connectivity of words in the child's *known vocabulary* and a model based only on the connectivity of words in the *language environment*. Here, a word is learned proportionally to node "importance" in the graph but conditioned on the child knowing at least one of the words that gives rise to the edge in the graph. We formalize this by defining the probability of learning word i as the centrality of that word if that word were added to the graph (indicated here as the union of the known words of a particular child K_x and word i). Words are learned if they are more central to the known vocabulary graph than other unknown words. For example, if a child's vocabulary network has an animal component and a water component, bridging words like duck and fish might have higher betweenness centrality when added to the graph than other unknown words. This model presumes that the words that are most likely to be learned are words that will become most important in the productive vocabulary graph (in comparison to other unknown words) once learned. We define word importance as

$$\delta_{i,x} \propto c_i^{(K_x \cup i)}, \tag{5}$$

when lure of the associates is the presumed model by which the network representations of small children grow.

Allowing for word importance to be based only on pairwise relationships in which at least one element of the pair is known suggests that children may need context and understanding to ground learning and allows lure of the associates to have a stronger relationship to the child's known vocabulary than preferential acquisition and a stronger relationship to the language environment than preferential growth.

For the current analysis, the way in which we define "word importance" is based on not only growth models and computation of δ's but how we define centrality or c in these models. Centrality measures, which are embedded in our calculation of growth values, calculate the role of each node in the graph. While there are many different types of centrality capturing different types of node importance, we consider three types of centrality that we believe are cognitively relevant and may capture some meaningful aspects of language acquisition in young children. The first is in-degree centrality as put forth and considered on normative vocabulary snapshots by Hills and colleagues [2]. We also consider undirected betweenness and closeness centralities. Though these measures are correlated, there are differences in terms of interpretation when using these centrality measures. *Degree* centrality models presume that the number of neighbors a word has is relevant to making a prediction of future word learning. Degree centrality is considered a more local measure of centrality as a node's degree centrality is based only on the node's immediate neighborhood and not the global location of the word in the full lexical network. *Betweenness* centrality instead suggests that words are more likely to be learned if they provide new and/or shorter paths between currently known words. *Closeness* centrality suggests that vocabularies may grow based on minimizing distance between words, even as the network grows in size. Both betweenness and closeness centrality are considered to be more global measures of centrality as changes in the global graph structure will influence the node-level centrality measures. On average, we can expect betweenness centrality to change more drastically than closeness when we change the graph structure.

3.2. Longitudinal CDI Data. To evaluate our models, we need detailed data of the words that a child learns through the course of development. As mentioned above, one well-established way to measure and characterize toddlers' lexicons is to use vocabulary checklists, such as the MacArthur-Bates Communicative Development Inventory (CDI) [17]. The CDI checklist, completed by parents, indicates whether or not the child *produces* each word of a fixed set of words. These parent-reported vocabulary measures have been shown to be effective in evaluating children's communicative skills up to 30 months of age [22, 23]. The CDI: Words and Sentences Toddler Form is a checklist of approximately 700 early words that are typically produced by the at least half of children by 30 months of age.

Longitudinal CDI data from 83 monolingual toddlers (37 females) were collected as part of a 12-month study, conducted at the University of Colorado Boulder, Colunga Lab. Recruitment for the study was done in three phases and was biased toward recruiting children that were learning language at a slower rate than their peers (classically called *late talkers*). Language ability was evaluated based on *CDI percentile*. CDI percentile is an estimate of the number of children one could expect who would have a vocabulary equal in size or smaller than the observed child's vocabulary when controlling for age and gender. Observed CDI percentiles, as computed based on CDI norming data, spanned all language learning levels with an average learning percentile of 37.3 at the first of 12 visits and 61.3 at the end. The mean age of children was 17.5 months (range 15.4–19.3) at the first visit. On average, we have 10.9 CDIs (minimum of 2 and maximum of 12) for each child. Altogether, we have a total of 908 CDI forms. Figure 2 represents the type of longitudinal data utilized for modeling. For modeling purposes, we consider the change in vocabulary, or the difference between two sequential CDIs from the same child, to be a *vocabulary snapshot*, with the first CDI being the initial CDI and the second being the prediction CDI. In total, we have 825 CDI vocabulary snapshots.

One goal of our modeling approach is to predict the individual words a child is likely to learn next. To this end, we model how the network of an individual language learner changes over time by predicting when nodes will enter the graph. Figure 3 visualizes a learning trajectory of a child via four network graphs that are changing over time based on the addition of nodes to the graph. Edges are based on the McRae feature norms [18]. Our research goal is to construct a network growth model that captures the *evolving networks structure* through accurately predicting the individual words the child is likely to learn next.

4. Experimental Validation

For training individualized network growth models, we utilize cross validation. Training data consist of 60% of the

	Age	Sex	...	Voc. size	Dog	House	...	Zoo
Child A	16.2	F	...	32	0	0	...	0
	17.1	F	...	49	1	0	...	0
	18.9	F	...	132	1	0	...	1
Child B	19.3	M	...	257	1	0	...	0
	20.5	M	...	345	1	1	...	0

Figure 2: Example longitudinal CDI data for two children. Our models are trying to predict when a word enters the graph or the grey cells.

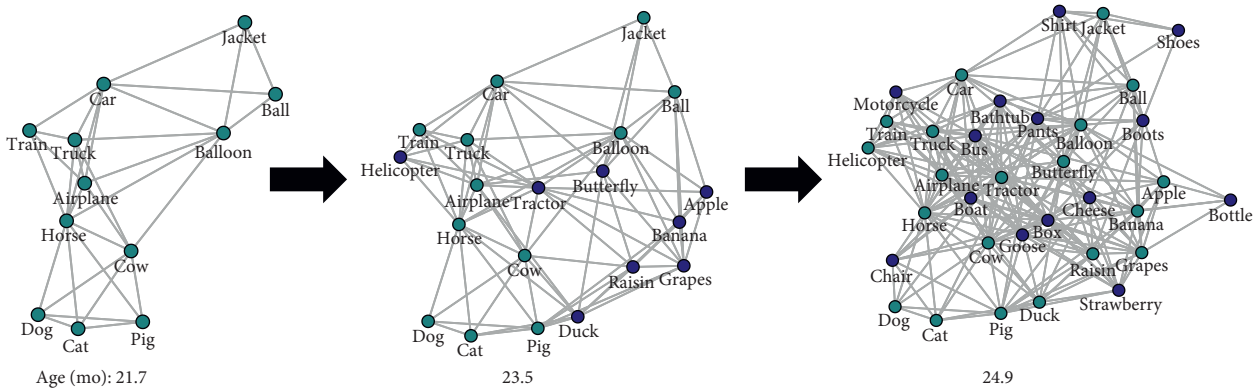


Figure 3: Network representation of a child's growing productive vocabulary.

children and a total of 484 snapshots. Validation and test sets each include 20% of children. All snapshots of a single child were included in only one group. Because of the variability in number of snapshots of an individual child, we verified that the size of the data (e.g., number of individual snapshots) also had similar proportions to the 60/20/20 split of children's data. Note that models were evaluated not only on performance of unseen snapshots but also on unseen children.

Each training example consisted of a *vocabulary snapshot* or a paired set of CDI reports collected at approximately one month intervals. The initial CDI was used to construct the child's known vocabulary network. The *growth value* was then calculated assuming a specific network representation, growth process, and centrality, conditioned on the child's current productive vocabulary where relevant (see equation (3) through (5) for specifications on how the growth values were calculated for the three different growth processes). The growth value is then combined with a baserate learning value from the CDI norms; these inputs are then converted to a probability through a logistic transformation as discussed in equation (1), with binary threshold θ, scale, and intercept parameters optimized and validated on training and validation sets, respectively. The resulting probability indicates, for each word not known by the child at the initial CDI, the probability that the word will be learned by the next CDI for that child. Observations are approximately a month apart, but the time between snapshots varies slightly across observations due to difficulties of scheduling.

Each model has a total of four parameters which we review in detail here. A threshold, converting the weighted graph to a binary graph, is optimized. We threshold the graph because the binary representation aids in interpretability of results and provides quicker convergence and more stability to the models. We note that this is a simplifying assumption made out of computational convenience. Although the acquisition of a word is generally a protracted process, with children remembering, forgetting, and refining the referential scope of a word over a period of time [24, 25], the data with which we have to build these networks are a binary judgment on the part of the parents (whether the child says or does not say the word), and we have no direct access to the child's possible specific representations of the word. In addition, as we are exploring different lexical representations and attempting to learn something from comparing across them, optimizing the threshold for each representation seems like a fair thing to do. Finally, the use of a binary network makes structural comparisons across networks more interpretable. The threshold can be thought of as a means to cancel out distributional and measurement noise and to instead highlight connections and relationships that are strong enough to garner attention of a child rather than a weighted network which would provide a notion of the importance of relationships between objects. With this binary graph and the child's known vocabulary at the initial CDI of our snapshot, we then compute a baseline probability $\delta_{\text{cdi}_{(i,x)}}$ and the delta value based on the network growth model as defined in Section 3.1. We then fit a logistic regression, converting these growth values and CDI baseline values to probabilities. Note that we are fitting both the network growth value and the growth value from the CDI age of acquisition norms simultaneously and for each network representation. This is used to account for the fact that some words are learned based on developmental rather than linguistic trajectories. We repeat our optimization procedure using expectation maximization on the threshold to convert the weighted graph to a binary graph,

probabilistically selecting a threshold and learning the optimal logistic regression parameters and network threshold based on optimal negative log-likelihood values on a validation set. Because the network size varies based on the network representation, we assume words on the CDI, but not in the network representation, are learned according to the normative, age-specific acquisition baseline model, e.g., equation (1) with β_2 fixed to 0.

Model selection is based on minimizing negative log-likelihoods for predictions to unknown words in the validation set. This measure penalizes both overestimation and underestimation of learning specific words such that if the model is highly confident a word is not learned and the word is learned; this miscalculation contributes equally to the error of confidence that a word is learned when it is not. We also note that, because we are only including predictions to words that are unknown at the initial CDI in the snapshot, this measure overrepresents children who have small vocabularies because they have many more CDI words that they could possibly learn. Using the validation data, we compare model performance to the age of acquisition baseline model. Because we are exploring network representation, network growth mechanisms, and centrality, we select a subset of the most predictive models to extend to the test set and to discuss in terms of insights into lexical acquisition.

5. Results and Discussion from Network Growth Framework

5.1. Macro Level: Effect of Network Structure. We begin by characterizing the structure of the network representations on the maximal lexicon. To create the network representation, we take the overlapping words between network representation and the CDI norms. We consider words and their variants, like shoe and shoes, to be equivalent and, when necessary, average their representations. Alignment between CDI words and network representations results in networks of different sizes ranging from 133 words to the full 677 words included in our longitudinal CDI assessment. In Table 1, we compare the structure of the observed networks to two types of random models. We note that the original network representations are weighted. We use the threshold as learned by the best performing growth model and centrality when characterizing the network structure here (see Table 2 for the threshold values). We compare the structure of the lexicon to two random models to better understand how representation of these language graphs differs from randomly generated semantic graphs.

The first comparison of the observed language graph to randomly generated, size-matched networks is the configuration model. The configuration model results in the construction of random network variations where the degree distribution matches the degree distribution of the observed network. We abbreviate this model in Table 1 as CM. We also generate a random model via a variant of preferential attachment [8] as proposed in the language acquisition paper of Steyvers and Tenenbaum [1]. We abbreviate this model as ST. In the Steyvers and Tenenbaum model, the network starts with a few seed words, and then at each iteration, an attachment word is selected proportional to the word's (current) degree. Once an attachment point is selected, a new node is added, with an edge between it and the attachment point. Then, the new node is connected to neighbors of the attachment point with some probability $\alpha = 0.9$. Neighbors are sampled until the new node has a degree equal to the mean degree or until every node in the current vocabulary is considered (see [1] for more information on this model). Because of the iterative edge building of the ST model, this model does not always converge on dense graphs since there are not enough available neighbors to maintain the high observed degree distribution at early iterations; thus, we do not include the results for the ST model on phonology as the resulting random network deviates greatly from the observed network even on network density. The lower density of the ST model comes from the fact that the observed graph has a density near 1 and edges added early cannot maintain the overall network density of the observed phonological network. We present the size of the graph, the density, the average degree, transitivity (e.g., the probability that a is connected to c given a is connected to b and b is connected to c), mean geodesic, graph diameter or maximal shortest path, and the assortativity coefficient for (1) our observed networks, (2) configuration models (CM), and (3) Steyvers and Tenenbaum preferential attachment models (ST). We average over 100 runs of each random model in Table 1 and report mean estimates as well as standard deviations around these estimates.

When comparing the observed early semantic graphs to those generated by the random variants, we find that the local structure and assortativity of the observed networks cannot be well captured by the random graphs we considered. The phonological network representation is difficult to compare since the average node degree is almost the size of the full graph, and thus we focus instead on the semantic networks. We find that our observed networks have more local structure as captured by transitivity than would be expected by the degree distribution alone (in comparison to CM). Looking at the ST model, we find that the process of adding edges results in a slightly less dense network as nodes added early cannot form edges with enough neighbors. The transitivity measure of the networks generated by this model is close to the observed value in our semantic networks due to the fact that edges are added with high probability and between direct neighbors of the attachment node. Both random models, however, fail to recreate the geodesic distance of these observed networks, underestimating the distance. Both models also fail to recreate the assortativity observed in the language networks, and in two cases, the random models reverse the direction of this assortativity.

Although Table 1 aims to summarize the network representation as a whole, we use an individual child's specific productive vocabulary to induce a network which our growth model uses to predict learning of unknown words. To better understand the structure of these lexical networks, we consider the specific network structure of our snapshot data. In Figure 4, we plot the density, average local clustering

Table 1: Network summary statistics of network size (included next to the observed network name and consistent across all random models), network density, mean degree ($\langle k \rangle$), measure of transitivity (trans.), mean geodesic distance (geodist.), diameter (diam.), and assortativity coefficients (assort).

Network	Density	$\langle k \rangle$	Trans.	Geodist.	Diam.	Assort.
McRae (133)	0.191	25.33	0.524	1.99	4	−0.037
McRae CM	0.191 ($2e-3$)	25.33 (0.36)	0.309 ($5e-3$)	1.86 (0.01)	4.1 (0.31)	0.016 (0.017)
McRae ST	0.187 ($1e-3$)	24.81 (0.02)	0.532 ($7e-3$)	1.85 (0.01)	3.9 (0.30)	−0.142 (0.011)
Nelson (545)	$9.8e-3$	5.38	0.153	4.65	12	0.012
Nelson CM	$9.7e-3$ ($1e-4$)	5.36 (0.07)	0.027 ($1e-3$)	4.04 (0.04)	9.8 (0.78)	−0.004 (0.018)
Nelson ST	$9.1e-3$ ($8e-5$)	4.99 (0.01)	0.109 ($6e-3$)	3.07 (0.05)	6.5 (0.31)	−0.137 (0.013)
Phono (677)	0.956	646.34	0.987	1.04	3	0.013
Phono CM	0.952 ($2e-4$)	643.67 (0.12)	0.983 ($1e-4$)	1.04 ($1e-4$)	3 (0)	−0.004 ($7e-4$)

Note. We consider our observed networks, as well as networks constructed via configuration modeling (CM) networks (in which the degree distribution of the observed network is matched), and a preferential attachment algorithm applied to early acquisition (ST) as discussed in Steyvers and Tenenbaum's 2005 paper [1]. It is easy to see that in most cases, the configuration model cannot capture transitivity or the local structure and no random model can account for the assortativity coefficients.

Table 2: Best performing models on each of the network representations.

Network	Model	Cent	Thresh	Density	llk	CDI llk	β_{cdi}	β_{δ}
McRae	Lure	Close	0.091	0.192	0.402	0.432	0.143	0.759
Nelson	Lure	Degree	0.011	0.010	0.384	0.394	0.651	0.357
Phono	Lure	Close	0.101	0.956	0.364	0.380	0.377	0.672

Note. All models reliably outperform random (CDI llk) when applied to validation data. The network threshold (thresh) and network density as well as the log-likelihood of the model and the baseline CDI model are reported. The normalized logistic transform of the network (β_{δ}) and CDI norms (β_{cdi}) is reported for comparison of the importance of the CDI norms versus the network word features.

coefficient, % of vocabulary in the giant component, and the assortativity coefficient of the vocabulary network of these developing network lexicons. Sorting by the child's vocabulary size (top) and percentile (bottom), even though these measures are highly correlated, we find structural differences in our networks and across the language learning trajectories of children. As children learn language, their vocabularies necessarily get larger (top) however we are also interested in how a child's language learning ability might influence the emergence of this structure. For example, it is possible that children in the highest percentiles have a much more dense network than children with low percentiles, and thus we might be able to experimentally test whether increasing density of these language graphs (by teaching children words selected by adding nodes via a mechanism like lure of the associates) could actually influence language learning in toddlers. The bottom frame of Figure 4 aims to visualize these results. Note that some of the lines indicate that not all measures vary according to vocabulary size or percentile. For example, even though assortativity was a unique feature of the language networks as compared to randomly generated networks, this measure does not vary in relation to development as seen by the nearly flat line in column 4 of Figure 4. We also include all CDI snapshots in this figure to highlight variability of these network measures across children.

With the eventual goal of intervention applications, children with low percentiles (usually those with percentile's less than 20%) are of specific interest. One challenge in dealing with these early delays is that about half of children who are younger than two and in these lower percentiles go one to catch up to their peers without any intervention, whereas the rest show persistent delays [26]. Being able to identify those children most in need of intervention at an early age would be critical to effectively alter the developmental trajectory. Understanding what might be contributing to children learning language at a slower pace than their peers is complicated by the fact that small vocabularies have relatively high variability when measured with CDI checklists as there are more possible words that could be learned.

However, looking at network representations of these snapshots of children with small vocabularies and children with low CDI percentiles, we see stark differences in network structure, providing a promising direction for future work looking at diagnostics and interventions on late talking children. When comparing the best fit smoothed average (local polynomial regression fitting) of the left most data across the top and bottom rows, we find that children who have lower percentiles have larger giant components than younger children with similar vocabulary size in the case of the Nelson and McRae network representation, but that this giant component size becomes connected to all words in the network later in development. In the case of the Nelson network, we find that local structure as measured by transitivity is much higher in late talkers than in children with small vocabularies, even though late talkers often have small vocabularies. We also find that density and transitivity are higher for late talkers than younger children in the phonological network representation as well. These differences could suggest that late talkers and/or younger children may learn differently and will be a focus of our future research.

5.2. Mezzo Level: Effect of Network Growth. Turing to network growth models, we consider the role of the network representation, growth process, and node importance or

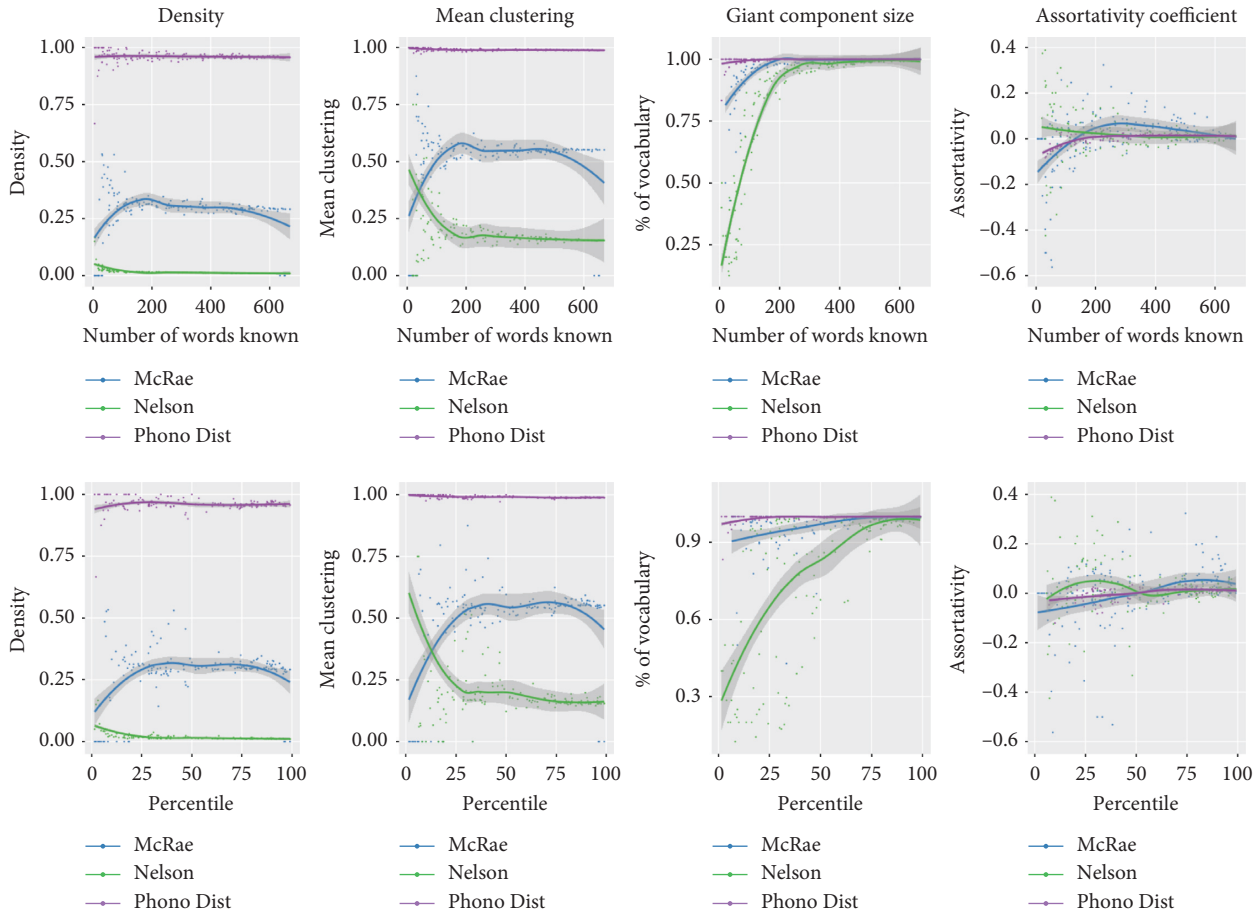


FIGURE 4: Plots of network measures as a function of the child's vocabulary size (top) or percentile (bottom) for validation snapshots. Different network representations (colored) indicate that the snapshot structure varies over the course of development and dependent on language ability.

utility. In total, we consider 27 models (3 network representations by 3 growth models by 3 centrality models) in addition to a model that includes only the delta values from the *CDI norms* which assumes that the network growth model contributes no information to the model. Because the network representations include a different number of words and we would expect better performance of our network models on words that are in the graph, there is a unique CDI norm baseline for each representation. We can also compare across representations by using the CDI norm baseline for words that are not in the specific network graph. We find this type of analysis masks much of the impact of the network representations especially because the McRae network has only 133 words. We thus neglect this comparison for the remainder of the paper.

We begin by comparing performance of all our models to the word-matched CDI model. Figure 5 plots the improvement of model performance for each network representation aggregated by growth mechanism (Figure 5(a)) and centrality measures (Figure 5(b)); zero indicates performance of the baseline CDI model. The y-axis indicates improvement in percent of log-likelihood error over our baseline model. The x-axis is organized by network representation, but the specific x-values within a network representation are not meaningful and are only used for readability. Positive values in Figure 5 indicate the model outperforms the network-specific baseline model. Note that a specific model can perform worse than, or near to, the baseline because here we are considering performance on the validation data. Performance near zero may suggest overfitting or high parameter sensitivity.

One clear result that can be seen from this figure is that there is not a single dominating growth mechanism or centrality. If this were the case across all network representations, we would expect clear line separation of the models in terms of growth mechanism or centrality and a similar color gradient across the three network representations. We in fact only see clear separation in the case of the Nelson model where *Lure of the Associates* is the only model that shows a reliable improvement over all other models we consider (middle left panel). Even considering the Nelson representation, we still see a substantial interaction of centrality within the growth process of lure of the associates (middle right panel). The fact that the graphs are not easily separable along growth mechanism (Figure 5(a)) or centrality (Figure 5(b)) strongly suggests an important interaction between the mezzo (growth process) and micro (node-level interaction) levels of analysis in the domain of modeling acquisition.

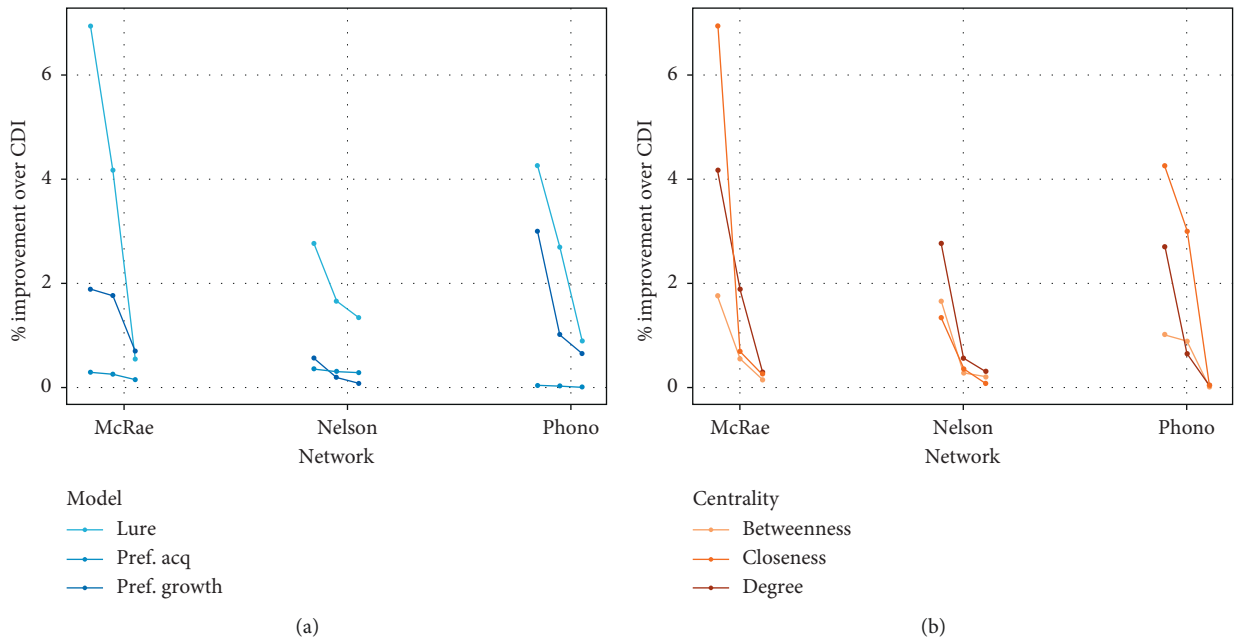


Figure 5: Percent improvement in log-likelihood of predictions on validation as compared to CDI baseline model. Performance is clustered by growth mechanism (a) or centrality (b). Positive y-values indicate improvement over the CDI baseline model; position along the x-axis is for better legibility and not reflective of performance across models.

In the three network representations we have chosen for our current analysis, all models perform at or above chance. However, Figure 5 suggests that there is no clear main effect of growth mechanism or centrality that is consistent across all network representations, or even within a network representation. Thus, instead of evaluating the significance of each model, we choose to focus only on the best model for each network representation. We summarize these models in Table 2. We report performance of the model in terms of negative log-likelihood and include the performance of the CDI baseline model on the specific subset of words included in each network representation as well such that the reader can understand the improvement in terms of average negative log-likelihood of the network model as compared to the baseline CDI only model. Here we find that *Lure of the Associates* is the best performing growth mechanism regardless of network representation. The fact that *Lure of the Associates* is the best growth process supports the idea that, when predicting future acquisition, we should consider both the child's existing lexical knowledge as well as the structure of language in the child's environment.

In considering the mezzo level, beyond finding that *Lure of the Associates* is the best performing model, rarely do we see *Preferential Acquisition* significantly outperforming the baseline model, suggesting that it is not a useful mechanism for predicting the vocabulary growth of an *individual child*. Recall that *Preferential Acquisition* assumes that words are learned proportionally to their centrality in the full vocabulary graph. Previous results found this mechanism (using degree centrality) to be the most accurate model when accounting for normative acquisition trends [2] (we replicate their results using our models on normative acquisition, confirming their finding that preferential acquisition is the best performing model for normative acquisition). The failure here to account for individual vocabulary growth is likely because this model does not adapt well to individual differences and, in fact, does not adjust predictions to a child's vocabulary knowledge in any meaningful way. The inability of this model to predict individual acquisition trajectories and the success of *Lure of the Associates* suggests two main findings: (1) normative acquisition is quite different than the acquisition of any particular child and (2) the content of the child's vocabulary is important and predictive of which words a child is likely to learn next. *The Lure of the Associates* increased performance over preferential growth may also suggest that children need grounding in their productive vocabulary for connectivity in the lexical network to aid in learning. This poses an interesting direction for future work on interventions related to accelerating vocabulary acquisition.

5.3. Micro Level: Effect of Centrality. Analyzing the role of network centrality suggests that global centrality measures (e.g., determined and influenced by the structure of the full graph) are as accurate as local measures (e.g., determined only by local neighbors). In the case of global measures, the addition of a node may change the full centrality of the network much more drastically than in the case of the local centrality measures, but this potentially large change does not seem to aid the predictive power of these growth models. The most local centrality measure we consider is *in-degree* as this measure only considers immediate neighbors whereas the more global measures are closeness and *betweenness* as these measures include contributions from all nodes in the graph. Global centrality measures particularly affect *Lure of*

the Associates, as this model selects words for learning that most substantially increase the connectivity of the known vocabulary structure. While we see that closeness centrality is the most predictive in two out of the three growth processes we consider, we do not see a clear effect of centrality. It seems there is a complex interaction between growth model and centrality that requires further investigation in future work. Another possibility is that the performance of these models could interact with aspects of the learner such that a particular growth process or centrality is most predictive for certain types of learners. We explore this possibility below.

Even while our results do not strongly indicate a specific and clear effect of centrality, our results confirm that the chosen definition at each layer of analysis (micro, mezzo, and macro) does affect predictive performance. We additionally see that the growth processes have a substantial effect on our ability to generalize to unseen language trajectories and that, for the purposes of individual modeling, certain models are not flexible enough to predict the lexical acquisition of individual children above the predictive accuracy of CDI baseline. Finally, we show evidence for microeffects emerging from the way we choose to operationalize node importance. Taken collectively, these results indicate that consideration of the different levels of analysis—and importantly, of the interactions between levels—can help us capture and model the full complexity of the overall process of acquisition, while providing potentially novel hypotheses that can be tested further in modeling work as well as empirical studies on young children.

6. Predictive Accuracy on Unseen Children

For evaluation on the test data, we consider only the models in Table 2 to avoid statistically biasing our sample because these models were most predictive when generalizing from training to validation data. Looking at Table 3, and specifically the coefficients of the models, we can see that the standardized (the input features have been standardized such that the features have the same scale and standard deviation; this allows one to directly compare β coefficients) β coefficients suggest that there is a significant contribution of the network-based node importance (β_δ) as well as the CDI-based δ values (β_{cdi}). We can also see that the mean likelihood scores across all models are lower when including the network-based δ measures. Note we did not perform any statistical verification that these models reliably outperformed the CDI baseline at this time as to not bias our assessment on the test dataset. When extending our models to the test set (the results of which can be seen in Table 3), we explore different types of questions than those questions used for model selection and perform statistical analysis to verify if the network representation improves prediction accuracy.

In model selection, we optimized over four different parameters with respect to the average negative log-likelihood over *all* predictions to unseen words. We use the validation data to evaluate parameter fit. We considered the network threshold (to convert the network representation from a weighted network to a binary network), the growth process, the network centrality for each network representation, and the β weights for the CDI and network-based growth values. Because the binary network representation, and thus the calculations of the δs for the network representation, is affected by the threshold, we optimize the network threshold independently for each growth process and centrality measure. Note that minimizing log-likelihood is a slightly skewed measure because we are only predicting the learning of words (e.g., if a word is already known by the child, we do not make a prediction as to whether the word stays learned). This means that when minimizing log-likelihood over the whole dataset, younger children and children who know fewer words are overly represented in this sample. Additionally, it can easily be argued that we should care more about words that are learned than about words that stay unknown. This discrepancy between model optimization and research questions provides an interesting additional evaluation to our selected models.

We thus introduce additional measures in order to evaluate our selected models. The first measure is the average negative log-likelihood across unseen snapshots and has a similar interpretation to the log-likelihood values previously reported. Using the trained models, we make predictions individually to each unseen snapshot and compute the log-likelihood on that snapshot. We then average across snapshots. We can then easily compare, via paired t-test, whether our specific model outperforms the CDI baseline model. We additionally compute the percent overlap ($k \cap k'$). In this calculation, we assume that we know how many words (k) a child learned from one snapshot to the next and then compute the percentage of those words that are in the top k' words as predicted by the model. Additionally, we include the area under the curve (AUC) of the receiver operating characteristic (ROC) measure that considers the false-positive rate and the true-positive rate as the threshold for converting probabilities from 0 to 1 changes. The AUC summarizes this plot by computing the area under the curve constructed via varying the threshold. Also included are accuracy, precision, and recall values, computed at the population level assuming the best threshold from the ROC calculation. The results can be seen in Table 3.

The results included in Table 3 show that our network growth models are capable of predicting future word acquisition of individual children with higher accuracy than our baseline models (as all models are significant under multiple comparison correction with $\alpha = 0.05$). We also see that the nature of the network representation can affect our ability to improve over the baseline model as indicated by the fact that some representations such as McRae result in a lower accuracy than the baseline CDI-only model. Similarly, the Nelson network model has a lower percent overlap than the CDI baseline model. This suggests that some models may be better at predicting acquisition of words (e.g., improvement in percent overlap), while other models may be better at predicting what words are least likely to be learned as well as those words that are learned (e.g., AUC or llk measures). These are important directions for future exploration as it is unclear whether teaching children words that they are about to learn, or words that they are unlikely to

Table 3: Evaluation of model performance on test data.

Model	$\|V\|$	Snap. llk	p value	$k \cap k'$	AUC	Accuracy	Precision.	Recall
McRae	133	0.478	0.018	34.1	0.689	0.609	0.225	0.702
M. CDI	133	0.516		32.5	0.738	0.679	0.220	0.619
Nelson	534	0.531	0.030	37.9	0.761	0.677	0.256	0.726
N. CDI	534	0.561		38.5	0.764	0.691	0.270	0.668
Phono	677	0.460	0.008	34.6	0.734	0.690	0.238	0.627
P. CDI	677	0.502		31.6	0.767	0.697	0.207	0.664

Note. We include the size of the network representation (or number of words the model makes predictions for), the average negative log-likelihood for each unseen snapshot. We also report the p value of an F-test for nested models comparing the CDI-only model to the model containing word importance based on the network growth process. We include the overlap between the words learned by the child and the top words as predicted by the model or percent overlap ($k \cap k'$). Considering all predictions, we report the area under the curve (AUC) using ROC, accuracy, precision, and recall.

learn, or something in between, is the best means of intervention for long-term changes in the *rate* at which children learn words.

We also find strong evidence for an influence of the child's current vocabulary on the accuracy of our predictions as captured by the improved performance of the *Lure of the Associates* over the model that uses only information related to normative acquisition rates for words. These results suggest that modeling language acquisition benefits from the inclusion of interactions among the specific words a child knows. This finding highlights the importance of looking at the developing lexicon as a system, one that is likely to follow different developmental paths in different individuals. Additionally, we take these results as a promising validation of the cognitive insight and predictive modeling capabilities of a united network modeling framework that incorporates macro, mezzo, and micro levels of analysis.

7. Modeling Development

We are interested in a developmental perspective, beyond simply predicting future language learning. It is possible that these models are more accurate during certain periods in development or for earlier/later learned words. We note that this analysis is post hoc and that models were neither trained nor optimized to capture developmental effects. To consider these developmental effects, we take a naive approach of ordering the predictions by specific features we believe might be relevant for learning. Because each model has an individual baseline model optimized to predict the lexical items in a particular network representation, we compute the difference between the baseline CDI model for each representation and the network predictions such that values greater than zero indicate that the network model outperforms the baseline model *for that specific snapshot*. Although we showed above that all network models statistically outperform the baseline models, this does not mean that there is a performance boost for all children equally across development. Thus, we aggregate the individual likelihood predictions across a theoretical ordering of when words are learned (age of acquisition effects, e.g., certain words are on average learned earlier) or across snapshots (for developmental effects) to uncover trends in predictive accuracy of our network models.

Plotted in Figure 6, we show performance differences as compared to the CDI age of acquisition baseline. We construct orders based on the (1) average age a word is learned (AoA, average age of acquisition) as calculated based on normative acquisition trends (Figure 6(a)), (2) age of the child at time of prediction (Figure 6(b)), (3) CDI percentile of the child at the initial CDI of the snapshot (Figure 6(c)), and (4) vocabulary size of the child when the first CDI is collected (Figure 6(d)) Plotted points are the differences between the CDI baseline and the network model, colored based on network representation. We then fit a local polynomial regression which plots the smoothed locally weighted averages as a line to indicate performance of the models with respect to certain features of the child or vocabulary that we find interesting. We first note that when considering the word level, there are no clear trends as to when the network model outperforms the baseline model. However, when we organize the data based on child features, such as the age of the child at prediction, the smoothed curve of the log-likelihood difference suggests that certain representations perform better than others at different periods of development. Because age, percentile, and vocabulary size are all intercorrelated, we would expect the fitted lines to show similar trends. We find that the CDI age of acquisition baseline model outperforms our network models for some points in development; however, our network models are still significantly outperforming their corresponding CDI baseline models (see Table 3). We note that these results are only suggestive as there is a lot of noise in the data and this is a post hoc analysis. However, it suggests interesting trends that may be worthy of further investigation.

Performance indicates that the phonological network representation outperforms the baseline model, particularly early in development and toward the end of development. The fact that the phonological representation is most useful early in development also aligns with intuition that early in language learning, a child is limited by which words they can articulate intelligibly as much, if not more, than which words they know the meaning of. The improvement in predictive accuracy toward the end of development will take further investigation and more targeted theories as to the role of phonology later in the acquisition process. Another period of development in which our network models outperform baseline is for older children, children with high percentiles, and children with large vocabulary size. If a child knows nearly all the words, the network model can use information about the structure of those unknown words to pick up on statistically meaningful relationships, whereas the CDI

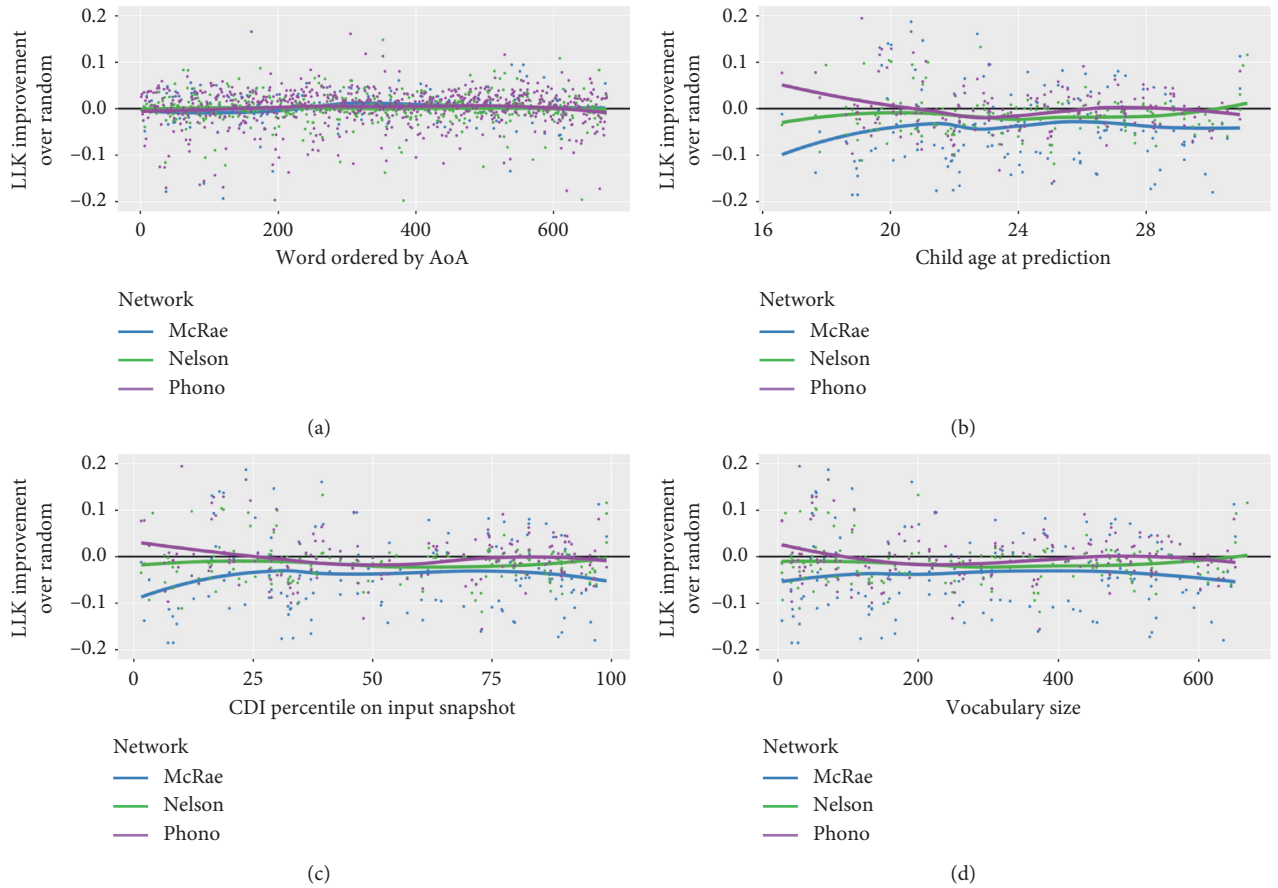


Figure 6: We consider performance of the best performing network growth models when sorted by average age of acquisition (AoA) (a) the child's age (b), CDI percentile (c), and vocabulary size (d). The lines indicate a locally weighted smoothing of the difference between each model's individual CDI baseline and the network model fit. Positive values indicate systematic improvement over the baseline CDI model for the population of snapshots in our test data. The points are individual differences that are used to compute the smoothed line.

model must rely on global and aggregated trends across all learners, neglecting individual differences. This improvement for children who know more words is further evidence that the child's current vocabulary aids in future language learning and that our network models may be capturing this relationship in meaningful and predictive ways.

These results collectively support the idea that individual modeling may be useful in capturing learning and developmental trends specifically for children who are far from the sample average. We can leverage this to our advantage particularly because the population we are most interested in modeling, namely, children with a low CDI percentile, is better predicted by these network models than the CDI baseline model. We find that the phonological network models are better than the CDI baseline norms particularly for very small CDI vocabularies and for children who have low CDI percentiles. The phonological network model generally performs well when compared to the baseline model for children between 16 and 22 months, despite the fact that this is not the age range where we have the most data. There are different ways of interpreting this finding. For example, these networks are based on the words that a toddler produces, according to the toddler's parent. It is possible that a network based on the words the toddler understands would preferentially weight semantics instead. This, however, would not explain how the phonological network is particularly predictive as compared to the baseline in the children with small vocabularies and low CDI percentiles. It is possible that our phonological network model may be capturing something about the nature of the deficit underlying the delay in some of these children—that being able to produce specific combinations of sounds is at least part of what is holding back their language development and thus offers a possible avenue for intervention [27, 28].

The high amount of noise in our plots when we aggregate by child features could be because there is a lot of variance in the network's explainability or because we have yet to find the features of a child's lexicon or child's development that correlate with the range when our network representations are most useful in predicting future language learning. In future work, we aim to further investigate not only network growth models that are capable of modeling individual differences but also hope to gain a better understanding of when our network models will be useful versus when the baseline CDI model is equally reliable.

8. Discussion and Future Direction

The network-based approach to modeling acquisition provides a unifying framework for studying the complex process

of word learning, allowing researchers to investigate individual differences and predict future vocabulary growth of individual children. We find that the definition of the edges in a network, the assumed process of network growth, and the network measure chosen to operationalize importance of words dramatically affect our ability to predict future acquisition. The importance of the three levels of analysis can be seen in the varying fits across our combinations of models. Although we find evidence, at particular points in development, that one network representation is clearly more accurate at predicting lexical acquisition of young children, this is an area requiring future study—if we understand contexts in which these network models are most useful for prediction, we will be able to use these models to provide insight into diagnostics and interventions.

Taken collectively, interesting results emerge that may provide future directions of study and possible interventions. For example, the phonological edit distance (*Phono.*) network captures acquisition trajectories of younger children and children with lower percentiles. This is in line with findings that suggest that late talkers are disproportionately behind in their production skills (the words they say) rather than their comprehension skills (the words they understand) [29]. We also find that at certain points in development, the CDI baseline model is comparable in predictive accuracy to our network-based approach. This suggests the possibility that there are attentional changes during the course of learning and, potentially, that later talkers or younger children learn differently than their peers. Again, there is empirical evidence from behavioral tests with toddlers to suggest that this is the case [30–32] This type of network modeling framework may allow for us to not only model differences in these groups but to explain the process of acquisition that leads to these differences.

The results suggest that phonological and semantic features are both important and relevant to language learning. While phonological network structure varies greatly from semantic network structure and network model results have shown differences in acquisition in relation to phonological networks [20, 33, 34], both of these individual network representations are useful in predictive models of lexical acquisition. In the future, we hope to jointly consider the effects of these levels of analysis in predictive modeling and as a means to understand the process of language development, possibly by building ensemble models within this framework or by extending this approach to multiplex representations [35, 36].

Our results also challenge previous work in the domain of network-based approaches to modeling acquisition. In the case of normative language acquisition modeling, it has been shown that *Preferential Acquisition* outperforms models based on the child's current vocabulary knowledge [2]. Here, however, we found that a model of preferential acquisition is unable to account for language growth at the level of individual children. All accurate predictive models of individual child trajectories use information about the child's productive vocabulary. We found strong evidence that *Lure of the Associates*, a model that considers *both* information about the child's current lexical knowledge and the language learning environment of the child as captured by the network of the full language graph, is the most predictive model regardless of the network representation we choose. This provides strong support to the notion that individual differences in learning can be related to *both* the language learning environment of the child and differences in the way that the child learns language. These results also highlight the need for modeling at the level of the individual rather than at the aggregate level such as via the age of acquisition norms. Although our models are ambiguous about how a child learns a specific word, the words themselves may have important and useful cues as to which are the relevant features that guide future lexical acquisition. These language learning cues may be related to the physical and linguistic environment that the child is immersed in or even to the child's specific interests. Either way, our modeling results strongly suggest an important contribution of the known vocabulary on future vocabulary growth.

In the best fitting models for each network representation, we find an impact of centrality measures on accuracy of the network growth models. One interpretation of this finding is that both global structure and local information are important to the relationship between emerging language graph and future language learning. This interplay between local and global network structure may be especially important for correcting language learning delays such as those of *late talking* children. The role of global centrality measures (betweenness) and local ones (degree) suggests that instead of teaching just a few words to help get children back on track, we may need to alter the connectivity of the graph instead. For example, previous training studies [37, 38] have shown that teaching words for shape-based categories (like "spoon" or "ball") accelerates vocabulary growth in typically developing children. Our results suggest that rather than having a static list of words that might be good for toddlers to learn, a goal may instead be to achieve a certain type of connectivity, for example, strengthening a nascent cluster of concepts (e.g., animal names), or teaching a specific word (e.g., "chicken") that might serve to link two clusters or categories (e.g., animal and food word clusters). We also note that we only consider network centrality measures here, but there are other ways of quantifying a node's importance. In the future, we hope to combine centrality measures with other non-network-related measures such as frequency or concreteness. While network connectivity seems to be useful in modeling language acquisition, other measures may provide additional support.

We note that this work is not the first work to use computational models to explore individual differences in development. In fact, there is a great deal of work focusing on capturing the structure of the child's language environment with fewer simplifying assumptions of our preferential acquisition and lure of the associates model [39–41]. Much of these computational modeling results suggest that children may be learning distributional information directly from the environment and that learned distributional information can explain production and use [41, 42]. These models are an important part of the overall quest to understand the acquisition process and complement our work

by studying the language learning environment. Our work instead makes many simplifying assumptions about this learning environment but with the aim of capturing the learning process and the impact of the child's current vocabulary knowledge on the overall process of acquisition.

The complexity of these network modeling results underscores the depth of the challenge that comes with modeling individual acquisition trends. Although here we present a first step in building up an accurate predictive network growth model, much more work is needed to explain why certain models perform in disparate ways, for different words, and for different language learners. In the future, with more data and more sophisticated network models, we hope to capture language learning with higher accuracy. Improving accuracy will also allow for the development and investigation of mechanistic models that offer explanations as to why certain children follow a specific language acquisition trajectory. We are particularly encouraged by the strong improvement of the network model over the CDI age of acquisition baseline model specifically for individual children who are learning language at a slower rate. Critically, although the accuracy in predicting future acquisition is a meaningful benchmark, one benefit of network analysis models over machine learning models is that the former models imply a possible mechanism of learning. Achieving a mechanistic understanding of the forces that shape early language acquisition is crucial in finding ways to improve the vocabulary of those children who are at higher risk of persistent language delays. In the future, we hope this work and other similar approaches of modeling at the level of the individual can pave the way for diagnostic and intervention tools capable not only of predicting but also explaining individual acquisition trends.

Data Availability

The data are not publically available, please contact the first author for the code to run the computational models.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by an award from the John Merck Scholars Fund and by NICHD (grant R01 HD067315) to Eliana Colunga. The authors also thank Michael Mozer, Matt Jones, Aaron Clauset, Tamara Sumner, and Massimo Stella for their helpful discussion and ideas around this project.

References

[1] M. Steyvers and J. B. Tenenbaum, "The large-scale structure of semantic networks: statistical analyses and a model of semantic growth," *Cognitive Science*, vol. 29, no. 1, pp. 41–78, 2005.

[2] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Longitudinal analysis of early semantic networks," *Psychological Science*, vol. 20, no. 6, pp. 729–739, 2009.

[3] N. M. Beckage, L. B. Smith, and T. T. Hills, "Small worlds and semantic network growth in typical and late talkers," *PLoS One*, vol. 6, no. 5, Article ID e19348, 2011.

[4] N. M. Beckage and E. Colunga, "Language networks as models of cognition: understanding cognition through language," in *Towards a Theoretical Framework of Analyzing Complex Linguistic Networks*, A. Mehler, A. Lücking, S. Banisch, P. Blanchard, and B. Job, Eds., pp. 3–30, Springer, Berlin, Germany, 2016.

[5] J. Borge-Holthoefer and A. Arenas, "Semantic networks: structure and dynamics," *Entropy*, vol. 12, no. 5, pp. 1264–1302, 2010.

[6] A. Baronchelli, R. Ferrer-i-Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen, "Networks in cognitive science," *Trends in Cognitive Sciences*, vol. 17, no. 7, pp. 348–360, 2013.

[7] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Categorical structure among shared features in networks of early-learned nouns," *Cognition*, vol. 112, no. 3, pp. 381–396, 2009.

[8] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[9] C. Smith, S. Carey, and M. Wiser, "On differentiation: a case study of the development of the concepts of size, weight, and density," *Cognition*, vol. 21, no. 3, pp. 177–237, 1985.

[10] E. V. Clark, "Making use of pragmatic inferences in the acquisition of meaning," in *The Construction of Meaning*, D. Beaver, S. Kaufmann, B. Z. Clark, and L. Casillas, Eds., pp. 45–58, CSLI Publications, Stanford, CA, USA, 2002.

[11] S. R. Waxman and E. M. Leddon, *Early Word Learning and Conceptual Development: Everything Had a Name, and Each Name Gave Birth to a New Thought*, Wiley-Blackwell Handbook of Childhood Cognitive Development, Malden, MA, USA, 2011.

[12] J. R. Saffran, "Statistical language learning," *Current Directions in Psychological Science*, vol. 12, no. 4, pp. 110–114, 2003.

[13] C. Yu and L. B. Smith, "Rapid word learning under uncertainty via cross-situational statistics," *Psychological Science*, vol. 18, no. 5, pp. 414–420, 2007.

[14] M. C. Frank, N. D. Goodman, and J. B. Tenenbaum, "Using speakers' referential intentions to model early cross-situational word learning," *Psychological Science*, vol. 20, no. 5, pp. 578–585, 2009.

[15] T. T. Hills, J. Maouene, B. Riordan, and L. B. Smith, "The associative structure of language: contextual diversity in early word learning," *Journal of Memory and Language*, vol. 63, no. 3, pp. 259–273, 2010.

[16] N. M. Beckage, A. Aguilar, and E. Colunga, "Modeling lexical acquisition through networks," in *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, TX, USA, July 2015.

[17] L. Fenson, P. S. Dale, J. S. Reznick et al., "Variability in early communicative development," *Monographs of the Society for Research in Child Development*, vol. 59, no. 5, pp. 1–185, 1994.

[18] K. McRae, G. S. Cree, M. S. Seidenberg, and C. McNorgan, "Semantic feature production norms for a large set of living and nonliving things," *Behavior Research Methods*, vol. 37, no. 4, pp. 547–559, 2005.

[19] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber, "The University of South Florida free association, rhyme, and word fragment norms," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 3, pp. 402–407, 2004.

[20] M. S. Vitevitch, "What can graph theory tell us about word learning and lexical retrieval?," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 2, pp. 408–422, 2008.

[21] J. S. Adelman, G. D. A. Brown, and J. F. Quesada, "Contextual diversity, not word frequency, determines word-naming and lexical decision times," *Psychological Science*, vol. 17, no. 9, pp. 814–823, 2006.

[22] D. J. Thal, L. O'Hanlon, M. Clemmons, and L. Fralin, "Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment," *Journal of Speech, Language, and Hearing Research*, vol. 42, no. 2, pp. 482–496, 1999.

[23] R. I. Arriaga, L. Fenson, T. Cronan, and S. J. Pethick, "Scores on the MacArthur Communicative Development Inventory of children from lowand middle-income families," *Applied Psycholinguistics*, vol. 19, no. 2, pp. 209–223, 1998.

[24] J. S. Horst and L. K. Samuelson, "Fast mapping but poor retention by 24-month-old infants," *Infancy*, vol. 13, no. 2, pp. 128–157, 2008.

[25] P. Bloom, *How Children Learn the Meanings of Words*, The MIT Press, Cambridge, MA, USA, 2002.

[26] J. Heilmann, S. E. Weismer, J. Evans, and C. Hollar, "Utility of the MacArthur–Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers," *American Journal of Speech-Language Pathology*, vol. 14, no. 1, pp. 40–51, 2005.

[27] A. L. Williams and M. Elbert, "A prospective longitudinal study of phonological development in late talkers," *Language, Speech, and Hearing Services in Schools*, vol. 34, no. 2, pp. 138–153, 2003.

[28] L. Girolametto, P. S. Pearce, and E. Weitzman, "Effects of lexical intervention on the phonology of late talkers," *Journal of Speech, Language, and Hearing Research*, vol. 40, no. 2, pp. 338–348, 1997.

[29] C. Desmarais, A. Sylvestre, F. Meyer, I. Bairati, and N. Rouleau, "Systematic review of the literature on characteristics of late-talking toddlers," *International Journal of Language & Communication Disorders*, vol. 43, no. 4, pp. 361–389, 2008.

[30] C. E. Sims, S. M. Schilling, and E. Colunga, "Beyond modeling abstractions: learning nouns over developmental time in atypical populations and individuals," *Frontiers in Psychology*, vol. 4, 2013.

[31] E. Colunga and C. E. Sims, "Not only size matters: early-talker and late-talker vocabularies support different word-learning biases in babies and networks," *Cognitive Science*, vol. 41, no. S1, pp. 73–95, 2017.

[32] S. S. Jones, "Late talkers show no shape bias in a novel name extension task," *Developmental Science*, vol. 6, no. 5, pp. 477–483, 2003.

[33] T. M. Gruenenfelder and D. B. Pisoni, "The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 3, pp. 596–609, 2009.

[34] M. Stella and M. Brede, "Patterns in the English language: phonological networks, percolation and assembly models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 5, p. P05006, 2015.

[35] M. Stella, N. M. Beckage, and M. Brede, "Multiplex lexical networks reveal patterns in early word acquisition in children," *Scientific Reports*, vol. 7, no. 1, p. 46730, 2017.

[36] M. Stella, N. M. Beckage, M. Brede, and M. De Domenico, "Multiplex model of mental lexicon reveals explosive learning in humans," *Scientific Reports*, vol. 8, no. 1, p. 2259, 2018.

[37] L. B. Smith, S. S. Jones, B. Landau, L. Gershkoff-Stowe, and L. Samuelson, "Object name learning provides on-the-job training for attention," *Psychological Science*, vol. 13, no. 1, pp. 13–19, 2002.

[38] L. K. Samuelson, "Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds," *Developmental Psychology*, vol. 38, no. 6, pp. 1016–1037, 2002.

[39] B. MacWhinney, *The Childes Project: The database, Volume 2*, Psychology Press, New York NY, USA, 3rd edition, 2000.

[40] S. R. Howell, D. Jankowicz, and S. Becker, "A model of grounded language acquisition: sensorimotor features improve lexical and grammatical learning," *Journal of Memory and Language*, vol. 53, no. 2, pp. 258–276, 2005.

[41] B. Riordan and M. N. Jones, "Redundancy in perceptual and linguistic experience: comparing feature-based and distributional models of semantic representation," *Topics in Cognitive Science*, vol. 3, no. 2, pp. 303–345, 2011.

[42] S. M. McCauley and M. H. Christiansen, "Language learning as language use: a cross-linguistic model of child language development," *Psychological Review*, vol. 126, no. 1, pp. 1–51, 2019.

Hindawi
Complexity
Volume 2019, Article ID 2108423, 24 pages
https://doi.org/10.1155/2019/2108423

Review Article

Cognitive Network Science: A Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics

Cynthia S. Q. Siew,[1,2] **Dirk U. Wulff**,[3,4] **Nicole M. Beckage**,[5] **and Yoed N. Kenett**[6]

[1]*University of Warwick, UK*
[2]*National University of Singapore, Singapore*
[3]*University of Basel, Switzerland*
[4]*Max Planck Institute for Human Development, Germany*
[5]*University of Wisconsin, USA*
[6]*University of Pennsylvania, USA*

Correspondence should be addressed to Cynthia S. Q. Siew; cynthia@nus.edu.sg

Received 7 September 2018; Revised 9 April 2019; Accepted 16 April 2019; Published 17 June 2019

Academic Editor: Ana Meštrović

Network science provides a set of quantitative methods to investigate complex systems, including human cognition. Although cognitive theories in different domains are strongly based on a network perspective, the application of network science methodologies to quantitatively study cognition has so far been limited in scope. This review demonstrates how network science approaches have been applied to the study of human cognition and how network science can uniquely address and provide novel insight on important questions related to the complexity of cognitive systems and the processes that occur within those systems. Drawing on the literature in cognitive network science, with a focus on semantic and lexical networks, we argue three key points. (i) Network science provides a powerful quantitative approach to represent cognitive systems. (ii) The network science approach enables cognitive scientists to achieve a deeper understanding of human cognition by capturing how the structure, i.e., the underlying network, and processes operating on a network structure interact to produce behavioral phenomena. (iii) Network science provides a quantitative framework to model the dynamics of cognitive systems, operationalized as structural changes in cognitive systems on different timescales and resolutions. Finally, we highlight key milestones that the field of cognitive network science needs to achieve as it matures in order to provide continued insights into the nature of cognitive structures and processes.

1. Introduction

Networks are everywhere. The friends you interact with in real life and on social media form your social network. Webpages form a network that you navigate through when you browse the World Wide Web. The same holds for roads, train tracks, or waterways for navigation in the real world. Over the past two decades, an increasing number of studies have applied network science methodologies across diverse scientific fields to study complex systems (e.g., [1–3]). Complex systems involve multiple components that interact with each other to give rise to complex behavior. This includes the human brain and the cognitive processes it gives rise to, such as memory and language (e.g., [4–8]). Network science is based on mathematical graph theory and provides a set of powerful quantitative methods to investigate these systems as networks (e.g., [9]).

In recent years, network science has become a popular tool in the study of structures and dynamics at the neural level of the brain [3, 10]. Despite the rich potential of the methods, research in cognitive phenomena has applied these tools to a lesser extent. This review aims to discuss how network science approaches have been applied to the study of human cognition and how network science can uniquely address and shed novel light on important questions related to cognitive systems and the processes that occur within those systems.

In particular, we aim to establish the following three points by an in-depth discussion of the extant literature on cognitive network science. This review is organized into three sections, each of these addressing one of the three issues.

(1) Network Science Provides a Quantitative Approach to Represent Cognitive Systems. One important goal of cognitive science is to model cognitive structures, for instance, semantic memory—our memory for facts or events (e.g., [5, 11, 12]), and the mental lexicon—the part of long-term memory where lexical representations are stored (e.g., [8, 13, 14]). This goal of formalizing cognitive representations is reflected in the diversity of approaches that have been employed including symbolic approaches (e.g., [15]), connectionist or neural network approaches (e.g., [16–18]), and combinations of the two (e.g., [19, 20]). We argue that a network science approach can provide a powerful alternative framework for modeling and quantifying cognitive representations in diverse domains. Network science provides a suite of computational tools that allows the cognitive scientist to explicitly examine the structural properties of cognitive systems—something that can be difficult to achieve with, for instance, connectionist approaches where the structure of a cognitive system is obfuscated within a black box of "hidden" layers [21]. Section 2 focuses on cognitive representations and architectures, demonstrating how network science approaches can be used to represent and describe the structural properties of cognitive systems. This section also introduces the reader to some basic terminology of networks.

(2) Network Science Facilitates a Deeper Understanding of Human Cognition by Allowing the Researcher to Consider How Network Structure and the Processes Operating on the Network Structure Interact to Produce Behavioral Phenomena. Another strength of the network science approach is the ability to not only quantify aspects related to the structure of cognitive systems, but also to model the processes that operate within these systems. For instance, a model of memory retrieval typically requires two core components—a representation of memories and a process to retrieve them. Within a network approach, these components might be modeled, for instance, as a network representation that depicts semantic memory as a network of similar concepts and a random walk process that walks randomly across the semantic network, emitting a sequence of nodes that approximates the outputs of the retrieval process as implemented by the random walk that traverses the underlying network structure. Generally, the joint consideration of structure and process emerges naturally from a network approach and provides a parsimonious account of human behavior and cognition in domains such as semantic memory and lexical retrieval. Section 3 focuses on the processes that occur in cognitive and lexical networks. We highlight how a thorough understanding of cognitive processes requires close consideration of how the structure of the cognitive system interacts with processes to account for complex human behavior.

(3) Network Science Provides a Framework to Model Structural Changes in Cognitive Systems on Multiple Scales. Another area of research that cognitive scientists are deeply involved in concerns the development and decline of cognitive systems [8]. Research in areas such as language acquisition and cognitive aging has revealed that cognitive systems are dynamic and are sensitive to changes in the linguistic environment [22], accumulation of experience over time [23, 24], and deficits or other age-related decline in sensory processing [25]. Dynamic changes may occur over longer timescales, reflecting the long-term accumulation of experiences, and smaller timescales, reflecting the dynamic nature of semantic memory in response to different contexts and experimental tasks [26]. We show how network science methods can provide new ways of quantifying and modeling the dynamics in these areas. Section 4 focuses on the dynamics of cognitive networks and specifically discusses research focusing on the factors that lead to structural changes in the network, and thus in cognition and behavior, on multiple timescales.

Cognitive science has largely employed network science methodologies to study the relationships between words and concepts (e.g., [5, 11, 27, 28]), aside from applications to social relationships (e.g., [29]). While a wide range of cognitive constructs can be represented as a network, the review will primarily focus on research studying memory and language-related phenomena using networks. Nevertheless, we note that network science can be a valuable tool far beyond the study of words and concepts (see Table 1) and encourage researchers to consider how network science methods can be used to address a broad spectrum of research questions in the cognitive sciences.

2. Cognitive Constructs as Networks

Networks are composed of two elements: *nodes* that represent the conceptual entities of interest (e.g., persons, websites, or words) and *edges* that represent the relationships among those units (e.g., friendship, hyperlinks, or semantic similarity). While additional aspects can be considered as is done in bipartite and multiplex networks (defined below), identifying these two basic elements in the system of study is sufficient to formalize the system as a network and to employ the powerful tools provided by network science. Network science approaches often capitalize on the fact that relationships between nodes (i.e., edges) can be as important as the nodes themselves, if not more important. A first challenge in studying cognitive systems as networks is to represent these systems in a meaningful way in terms of nodes and edges.

2.1. Network Representations of Cognitive Systems. Cognitive science traditionally has a strong interest in words and concepts as the basis for thought, reasoning, and communication (e.g., [37]). Much research has been dedicated to studying the properties of words, such as their frequency in natural language, their valence, or their concreteness (e.g., [38]). While these efforts have been instrumental for predicting the behavior of human memory and lexical retrieval, researchers have also found that much can be gained by considering the *relationships* between words within a network representation. Consider, for instance, the free association task that requires participants to freely generate associative responses to a given

Table 1: Examples of cognitive networks and their cognitive application.

Cognitive Network	Nodes	Edges	Relevant research areas
Semantic network	Words	Semantic relationships, including free associations, shared features, taxonomic, cooccurrence, semantic roles	Language acquisition; cognitive aging; semantic priming; creativity/insight; cognitive search and navigation; semantic memory
Form similarity network	Words	Phonological or orthographic similarity	Lexical retrieval; production; speech errors; memory recall; word learning
Syntactic network	Words; phrases; sentences	Cooccurrence; parse trees; syntactic dependencies	Language acquisition; language evolution; syntactic learning
Concept network	Concepts; ideas	Cooccurrence; causal; feature similarity	Learning; memory; concept formation
Informational network	Shapes; pictures; any unit of information	Temporal cooccurrence; communication; transmission	Statistical learning of external structure; information transmission
Clinical, personality networks	Symptoms; personality traits; items on a questionnaire	Statistical relationship such as partial correlations; comorbidity	Clinical psychopathology; personality disorders
Social network	People	Friendship; followers on social media; face to face interactions	Collective problem solving; decision making; echo chambers; polarization

cue, i.e., responding with the words *dog*, *purr*, and *kitten* to the cue word *cat* [31, 39, 40]. These associative responses can be represented as a semantic network, for example, by defining nodes as words and edges as the associative strength between words, i.e., the likelihood that one word is named as an association to another word. Representing the data in this way has allowed research to observe that a word's degree, a popular node metric derived from a network (defined below), predicts how well words can be learned [28, 41].

Associative strength is one of many options to construct networks of words. Edges between words can also be constructed based on the number of shared features (e.g., the concepts *banana* and *cheese* are connected as they are both yellow in color; [42], see also [43]), their semantic relations, such as synonymy (e.g., *happy* and *joy* share similar meanings), hypernymy (e.g., *maple* is a *tree*), and meronymy (e.g., a *bird* has a *beak*; see [44]), their phonological similarity (i.e., words that sound similar are connected to each other; [32, 45, 46]), their orthographic similarity (i.e., words that have similar spellings are connected to each other; [47, 48]), their cooccurrences in naturalistic speech [49], language corpora statistics [50], and manually annotated syntactic dependency relationships ([51, 52]; see [53], for a review). Likewise, research has studied different linguistic units other than words by mapping letters [54, 55], syllables or segments [56], or entire documents [57] onto nodes. Furthermore, research has applied network science methods to examine the semantic [58], stylistic [59, 60], typological [61–63], and phonological [45, 64] structure of languages other than English. Figure 1 shows examples of cognitive networks.

The minimum requirement for representing a cognitive system as a network is to identify nodes and edges. However, there is much more information that can be represented. For instance, networks can be specified with multiple types of nodes to distinguish between cues and responses in an associative network, creating a *bipartite network* [36]. Similarly, networks can be specified with multiple types of edges, for instance, to represent both phonological and semantic similarity between words, creating a *multiplex network* [65, 66]. Finally, edges can be *weighted* and *directed* in order to reflect the strength and directionality of a relationship, respectively. Directionality and strengths have been used to account, for instance, for the fact that a cue word triggers a response at a higher rate than the other way around, such as the case with *dog* frequently cuing the response *bone* but *bone* only infrequently cuing *dog*, with higher associates to words like *skeleton* or *body* instead [31, 36].

The construction of networks leaves much freedom to the researcher and renders it possible to represent a wide variety of cognitive systems as networks. For instance, the emerging area of network psychometrics represents statistical relationships between personality traits or items in a symptom questionnaire as a network, seeking to better understand the causal structure of human personality and psychological disorders (see [67, 68] for reviews). This approach is rapidly being established in personality and clinical research as a fruitful alternative to traditional approaches that use latent variable modeling, which assumes the presence of a latent variable that accounts for psychological and personality disorders, whereas the network approach emphasizes the relationships (i.e., edges) between symptoms and the importance of considering the causal pathways that can lead to the emergence of a disorder (e.g., [69–72]). Moreover, networks have been used to represent the external environment that people are embedded in, such as their social network or the informational space that learners are exposed to. Emerging work is showing that quantifying such external structures could lead to new insights into a number of topics of deep

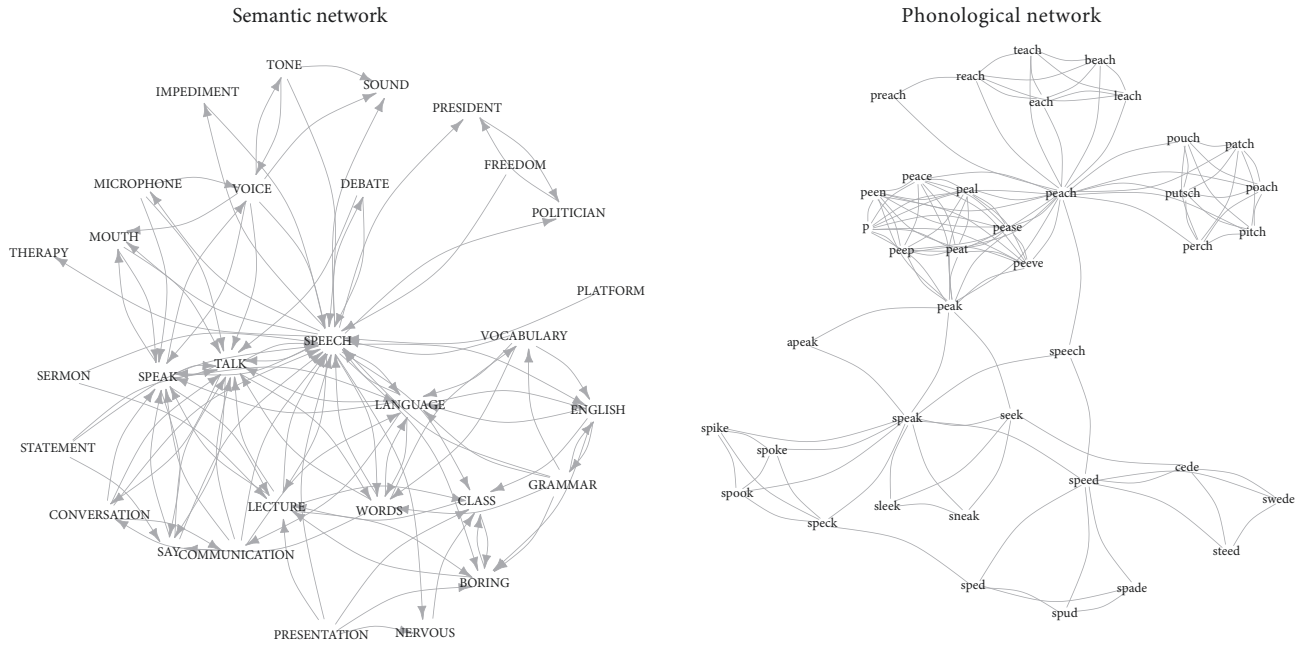


Figure 1: Examples of cognitive networks. Semantic network of free associations to the word *speech* (Left) and phonological network of words that sound similar to the word *speech* (Right).

interest to cognitive scientists, including the influence of a speaker's social network size on language processing [73], language evolution [74, 75], how students represent their conceptual understanding of a body of knowledge [76–78], problem solving and decision making in groups [79, 80], and how learners are able to extract the external structure of the world via statistical learning [81, 82]. See Table 1 for a summary of different types of cognitive networks and relevant cognitive science topics.

In summary, representing cognitive structures in terms of a network offers high degrees of flexibility to researchers investigating various cognitive phenomena (see Table 1). The nodes and edges in any cognitive network should represent theoretically motivated constructs, with nodes depicting an appropriate and relevant scale of representation and edges defining a meaningful relationship between nodes [83]. Choosing a network representation can be likened to choosing a measurement instrument. Different network representations will reveal different aspects of the underlying cognitive system. Ultimately, it is up to the researcher to decide which aspects to place the focus on.

2.2. How Can Network Structures Be Characterized? A key strength of studying cognitive systems as networks is the accessibility of reliable, well-established quantitative measures and tools, reflecting the long history of graph theory and its mathematical foundations [84], as well as its continual refinement and development (for instance, in the area of multiplex networks, [85]). In this section, we present common measures used to quantify aspects of networks at three scales of structure in network representations: (i) the microscopic structure, i.e., a "node's eye view" of structural properties of individual nodes and edges, (ii) the mesoscale, involving a subset of nodes and the substructures that they form, and (iii) the macroscale, i.e., a "network's eye view" summarizing the entire network structure. To highlight how these three different scales provide novel opportunities for the study of cognition, we review measures on each of these scales and how they have been employed to improve our understanding of cognition.

2.2.1. Microscopic Network Measures. At the microscopic level, network analysis examines different properties of nodes and edges, most commonly focusing on quantifying the "importance" of a node in the graph representation via measures of centrality [86–88].

One popular measure of node centrality is the node's *degree,* k_i, i.e., the number of edges connected to a node. Nodes of higher degrees are directly connected to a higher number of nodes in the network and can have an important role in, for instance, exchanging information across the network [4]. A node's degree also defines a node's "neighborhood size", a property that has often been used in the context of phonological and orthographic networks of word forms to predict lexical retrieval times, where edges commonly represent an edit distance of one phoneme [89] or one letter [90]. Here, the degree of a node defines the level of similarity of a word form to other word forms.

Another property of nodes derived from its immediate neighborhood is *local clustering coefficient*, c_i, which characterizes the extent to which the neighbors of a node are interconnected. Specifically, clustering coefficient measures the extent to which a node's neighbors are also neighbors of each other (similar to the notion of transitivity in social networks, i.e., are your friends also friends of each other; [91]). As shown in Figure 2, it is possible for a word with

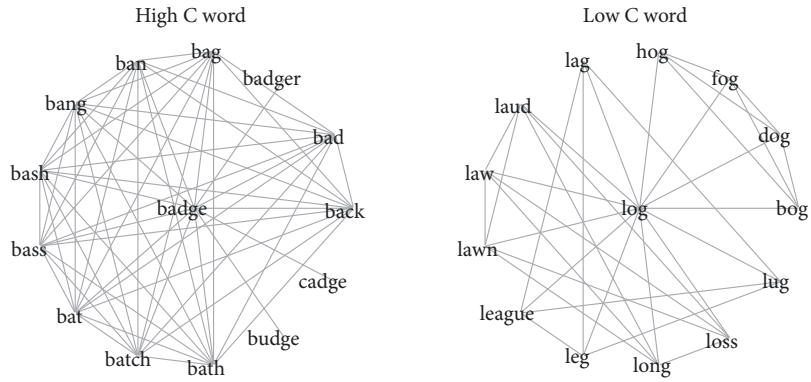


Figure 2: A word with high clustering coefficient (Left) and a word with low clustering coefficient (Right) are shown below. Notice that both words have the same number of phonological neighbors, i.e., degree. Adapted from Chan & Vitevitch [30].

the same degree to have different values of c_i, reflecting differences in the internal structure of their neighborhoods. Similar to a node's degree, a node's clustering coefficient will influence the flow of information through and around a node. Along these lines, recent psycholinguistic research has shown that people are sensitive to nuances in the similarity structure of words as operationalized by c_i. The clustering coefficient of words has small but measurable influences on how people recognize spoken [30] and written words [48, 92], produce speech [93], learn new words [94], and recall words in memory tasks [95].

Both node degree and local clustering coefficient consider only the immediate neighborhood of the node of interest. To characterize the importance of a node beyond its immediate neighborhood, measures such as closeness centrality or PageRank centrality can be used. *Closeness centrality* is computed as the inverse of the average shortest path length (defined in more detail below) to all other nodes in the network [96]. Closeness centrality of words computed from phonological and orthographic similarity networks has been shown to influence spoken [97] and visual word recognition [48], picture naming performance among people with aphasia [98], and performance in a mental navigation task [99]. *PageRank centrality* is widely known as an algorithm used to rank websites in Google search results [100]. PageRank centrality can be thought of in terms of a "fluid" that flows throughout the network and pools at the most important nodes. The general idea is that more important nodes (websites) receive more "fluid" (endorsement) from nodes (websites) that are themselves important in a recursive fashion. Although PageRank centrality was developed for the purpose of optimizing web search, Griffiths, Steyvers, and Firl [41] showed that an implementation of the PageRank algorithm on language networks was better able to account for people's responses in a fluency task as compared to traditional predictors such as word frequency or associate frequency—suggesting strong parallels between the mechanisms underlying successful information retrieval in search engines and in human memory.

Finally, the shortest path, or shortest distance between two nodes, $d(n_i, n_j)$, can be used to reveal something about the relationship of (nonneighboring) nodes rather than the nodes themselves. For instance, path length between two nodes in cognitive and language networks influence the misperception of spoken words [101], judgments of semantic relatedness [27, 102], and picture naming performance in people with aphasia [98].

Many other network measures that we do not discuss here in detail for the sake of brevity have been shown to have measurable effects on human behavior. These measures include assortative mixing by degree [103], key players [104, 105], whether a node resides in the largest connected component of the network or in smaller connected components [106, 107], network connectivity of a node's broader neighborhood that included its immediate neighbors and nonneighboring words [108], and many others that have yet to be thoroughly explored in the cognitive sciences, such as betweenness centrality (see [109], for a review detailing the influence of various network metrics on language processing, and [110], for a review of various centrality measures).

2.2.2. Mesoscopic Network Measures. At the mesoscopic level, research has focused on network community structure. Communities refer to the grouping of nodes into subnetworks based on their connectivity, i.e., how strongly they are interconnected. To identify communities in networks, existing algorithms aim to maximize connectivity within clusters while minimizing connectivity between clusters [111]. The identification of communities and the nodes they include can provide interesting insights regarding a cognitive system. In the domain of language, for instance, community detection can be used to identify semantic fields or categories [11]. The ability of community detection algorithms to describe a network in terms of a set of communities is typically captured using a measure known as modularity [111, 112]. Investigations at the neural level in the brain [113–115], for instance, have consistently shown how the community structure of neural networks changes with the progression of several different psychopathologies [116, 117].

Recent studies have also highlighted the significance of modularity in cognitive networks in both healthy and clinical populations [46, 118–120]. For example, the semantic network of individuals with high functioning autism (Asperger's syndrome) has been found to exhibit higher modularity

than matched controls, which offers one possible account of their rigidity in processing language [118]. Moreover, high modularity observed in the phonological network was suggested to constrain the spreading of activation in lexical retrieval [46] and higher modularity observed in semantic networks was negatively related to individual differences in creative ability [34, 121].

2.2.3. Macroscopic Network Measures. Measures of the macroscopic structure of networks speak to the overall organization of networks as opposed to node level or community level structures. These measures can reveal emergent properties of a system visible only when considering the network as a whole, which can play an important role in the system's behavior. Below we describe macroscopic network measures that have been used to study cognitive systems.

Average Node Measures. Networks are regularly characterized with the averages of local, node-based measures described above, such as average degree, average shortest path length, and average (local) clustering coefficient (e.g., [11, 28]). One pattern that frequently emerges in a variety of systems is known as a *small world structure*, characterized by high local clustering and moderate average shortest path lengths, relative to similarly sized, density matched, randomly drawn networks [91]. This small world property may be important in the domain of cognition for two reasons. First, small world structures have been found to be an almost universal property of real-world networks across diverse domains including biological networks (e.g., [122, 123]), social networks (e.g., [124]), and information networks (e.g., [125]). Second, small world structures may emerge from systematic growth processes that may adapt to environmental constraints to give rise to a beneficial structure. For instance, the small world structure of brain networks has been said to reflect the tradeoff between short neuronal distances between brain regions and the costs associated with creating these connections, with the conclusion that small world structure may provide a means to optimize the organizational structure of neurons [113]. The idea of optimized organization in brain networks is related to the idea of *network efficiency*, E_G, referring to a network's ability to quickly exchange information (Table 2; [126]).

"Small world-ness" is also a ubiquitous feature of many types of language networks, including semantic networks [28], phonological networks of various languages [32, 45], the orthographic network of English [48], and syntactic networks [52]; although we note that some people debate the usefulness of measuring small world structure as Watts and Strogatz [91] showed that even an extremely structured lattice will exhibit small world structure when a small amount of random rewiring of edges is introduced. Similar to the argument concerning brain networks, small world properties in language networks might arise due to two competing aspects of language learning and use—distinctiveness (e.g., each object having a unique word mapping) and memory constraints (e.g., the easiest language to learn is one where a single word refers to everything). These two competing features of language may result in the emergence of local clusters of similar meaning and form but a low average path length due to the influence of memory constraints resulting in the reuse of words and sublexical segments [50, 127]. Finally, the small world structure in semantic and lexical networks could provide important clues into how the structure of such cognitive systems might be exploited in order to maximize the efficiency of search processes within semantic memory or prevent catastrophic failures in lexical structures [11, 128].

Degree Distribution. The *degree distribution*, *P(k)*, of a network indicates how many nodes have a given number of connections in the network (i.e., its degree). In many naturally occurring networks many nodes have low degree (few connections) and a few nodes have very high degree (many connections). The degree distribution of some networks is often best approximated by a power law (see Clauset et al. [129] for a counter argument to the idea that power laws accurately capture degree distributions in networks.) such that $P(k) \approx k^{-\gamma}$ and is typically referred to as *scale-free* networks when the exponent, γ, is between 2 and 3 [129, 130]. The term scale-free refers to the fact that the second and higher order moments tend to go to infinity, implying that the degree distribution has infinite variance. Scale-free degree distributions include nodes with degrees much larger than the average degree of the network, which are often referred to as hubs. The scale-free property of networks has been linked to a network's resilience to random node failure. That is, studies of percolation processes have found that the connectivity in scale-free networks withstand a continued, random deletion of edges longer than networks with other degree distributions [131].

The degree distributions of semantic networks consistently follow an approximate power law distribution, with the exponents of the best fitting power laws converging at ~3 (see Figure 3, left; [28]). This has been demonstrated across semantic networks constructed from free associations [31] and from more complex semantic relationships (e.g., WordNet; [44]), suggesting commonalities in the semantic organization of word meanings ([132]; but see [133]). On the other hand, the degree distributions of phonological networks of various languages appear to be best fit by a truncated power law ([45]; see Figure 3, right), which have implications for candidate models of network growth.

To account for the ubiquity of scale-free degree distributions observed in naturally occurring networks, Barabási and Albert [134] proposed an influential model of network growth known as *preferential attachment* where, as new nodes are added to the network, these nodes are more likely to be connected to nodes with higher degree (i.e., more connections). Therefore, highly connected words are more likely to acquire new connections, resulting in a "rich-get-richer" effect. Steyvers and Tenenbaum [28] suggested that the growth of semantic networks could have occurred in a similar manner, by conceptualizing preferential attachment as a process of semantic differentiation in which words are likely to be learned if they connect to other words with many varied meanings, increasing the

Table 2: Definitions of network science terms and variables.

Term/variable	Definition
N	number of nodes, N, in graph
E	number of edges, E, in graph
network density	ratio of the number of edges to the maximum number of possible edges $\frac{2E}{N(N-1)}$
distance, $d(n_i, n_j)$	shortest path between node i and node j $d(n_i, n_j)$ where $n_i, n_j \in N$
average shortest path length, L	average length of shortest path between pairs of nodes $L = \frac{1}{N(N-1)} \bullet \sum_{i \neq j} d\left(n_i, n_j\right)$
diameter, D	largest shortest path between nodes $D = \max_{n_i \in N, n_j \in N} d\left(n_i, n_j\right)$
closeness centrality	inverse of the sum of the length of the shortest paths between node i and all other nodes in the graph $C_i = \frac{1}{\sum_j d\left(n_i, n_j\right)}$
degree, k_i	number of edges attached to node i
average degree, $\langle k \rangle$	average number of edges per node in network $\langle k \rangle = \frac{1}{N} \sum_{n=i}^{N} k_i$
local clustering coefficient, c_i	number of edges between the neighbors of node i divided by the maximum number of edges between those neighbors $c_i = \frac{2\lvert e_{jk} \rvert}{k_i(k_i - 1)}$ where $n_j, n_k \in N_i,\ e_{jk} \in E$
average clustering coefficient, $\langle C \rangle$	average clustering coefficient of nodes in the network $\langle C \rangle = \frac{1}{N} \sum_{n=i}^{N} c_i$
modularity, Q	proportion of edges that fall within subgroups of nodes minus the expected proportion if edges were randomly distributed, range $[-1, 1]$
average efficiency, E_G	measure of how efficiently information is exchanged in the network $E_G = \frac{1}{n(n-1)} \sum_{i \neq j \in N} \frac{1}{d\left(n_i, n_j\right)}$
largest connected component	largest group of nodes in the network that are connected to each other in a single component
degree distribution, $P(k)$	probability distribution of node degrees in the network
γ	power-law exponent for the degree distribution
Small world structure	network with short average path lengths and relatively high clustering coefficient (relative to a random graph with similar density)
scale-free network	network with a degree distribution that is power-law distributed

child's vocabulary and helping the child learn the various meanings of words. The truncated power law of phonological networks may instead suggest that a different underlying process operates on the growth of the phonological network.

2.3. Methodological Tools and Resources for Cognitive Network Analysis. In this section, we showcase a selection of resources that are available for cognitive scientists who are interested in estimating and analyzing cognitive networks. The number of available open source toolboxes to conduct cognitive network analysis is growing rapidly; below we cover a nonexhaustive list of the most relevant resources that enable the analysis, modeling, and visualization of cognitive and language networks.

NetworkToolbox: Methods and Measures for Brain, Cognitive, and Psychometric Network Analysis [135]: this R toolbox comprehensively implements network analysis and graph theory measures used in neuroscience, cognitive science, and psychology. NetworkToolbox includes various

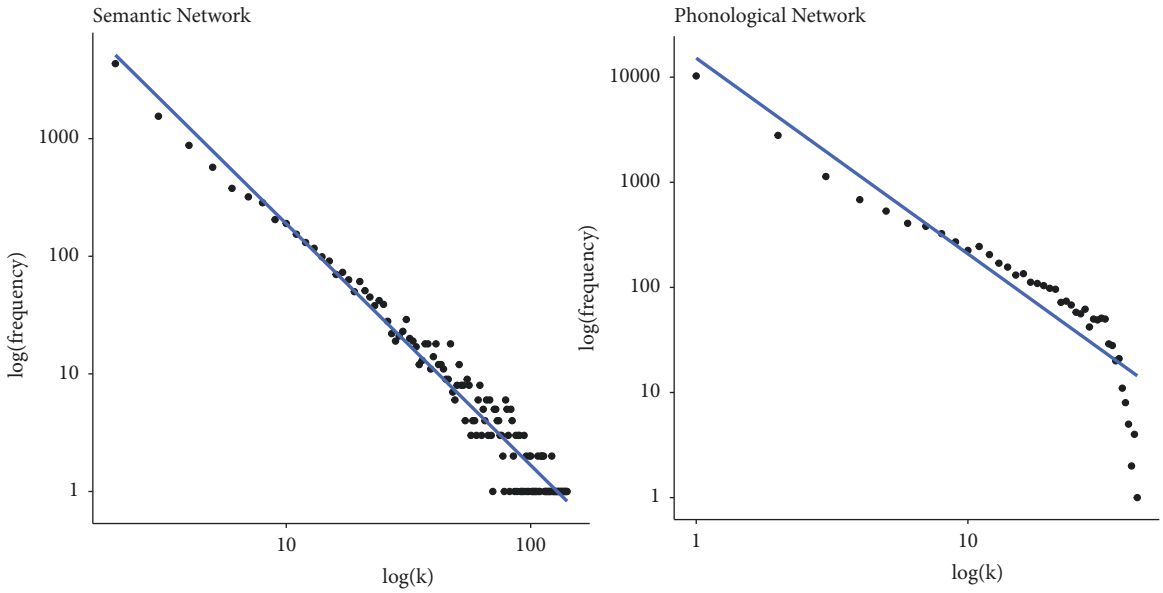


Figure 3: Degree distributions of semantic network constructed from Nelson et al.'s [31] free association norms and the phonological network from Vitevitch [32].

filtering methods for psychometric networks and efficiency-cost optimization for brain networks (https://cran.r-project.org/web/packages/NetworkToolbox/index.html).

SemNetCleaner: An Automated Cleaning Tool for Semantic and Linguistic Data [136]: this R package implements several functions that automatize the cleaning, binarizing, and spell-checking of text data. It also removes plurals and continuous strings, converges, and finalizes text data for semantic network analysis (https://cran.r-project.org/web/packages/SemNetCleaner/index.html).

memnet: Network Tools for Memory Research [24]: memnet provides a set of network science tools efficiently implemented for research in human memory. The memnet R package contains methods for inferring networks from verbal fluency data, implementing network growth models and diverse (switcher-) random walk processes, and tools to analyze and visualize networks (https://cran.r-project.org/web/packages/memnet/index.html).

spreadr: Simulating Spreading Activation in a Network [137]: spreadr enables cognitive scientists and psychologists to conduct computer simulations that implement spreading activation in a network representation (https://cran.r-project.org/web/packages/spreadr/index.html).

SNAFU: The Semantic Network and Fluency Utility [33]: SNAFU provides psychologists with the tools to generate semantic networks from category/verbal fluency data (e.g., name as many as animals as possible in 1 minute) and to compare the semantic networks of different groups or individuals (https://alab.psych.wisc.edu/projects/2017/12/08/snafu.html).

The Aging Lexicon Project [8]: the Aging Lexicon Project contains resources related to the study of the mental lexicon across the lifespan, including resources to measure and represent the linguistic environment, tools to quantify and model the mental lexicon over the lifespan, as well as a comprehensive list of open-access linguistic norms, natural language corpora, and behavioral megastudy data (https://aginglexicon.github.io/).

The Brain Connectivity Toolbox [138]: the brain connectivity toolbox is a comprehensive toolbox of MATLAB scripts, dedicated to analyzing networks. While several of the scripts compiled under this toolbox were primarily developed to analyze brain networks, they are well-suited to analyze cognitive networks as well (https://sites.google.com/site/bctnet/).

Psychometric network analysis: while outside the scope of this review, it is important to note that several R toolboxes have been developed to analyze psychometric networks (i.e., the analysis of psychometric questionnaires as networks to study psychopathology and personality). Such toolboxes include qgraph (https://cran.r-project.org/web/packages/qgraph/index.html), bootnet (https://cran.r-project.org/web/packages/bootnet/index.html), and mlVAR (https://cran.r-project.org/web/packages/mlVAR/index.html). For further detail, the reader is referred to Epskamp, Borsboom, and Fried [139].

2.4. Summary. This section provided an overview of various network measures at the micro-, meso-, and macrolevels and highlighted cases where these network measures were predictive of human behavior or provided novel insight into human cognition, predominantly drawn from the domain of language processing and semantic memory. Furthermore, we summarize a noncomprehensive list of main toolboxes that can allow the cognitive scientistto estimate and analyze cognitive networks.

3. Processes in Cognitive Networks

Now we turn to models that may account for processes that occur in the kinds of cognitive networks discussed

above. In this section we begin with an overview of the classic spreading activation theory in cognitive psychology and discuss extensions inspired by random walk models that are popular in network science research. We then discuss how these processes can account for behavioral findings implemented in a network representation and provide insight into latent cognitive processes from various domains in human cognition, including lexical retrieval, creativity, and cognitive search and navigation.

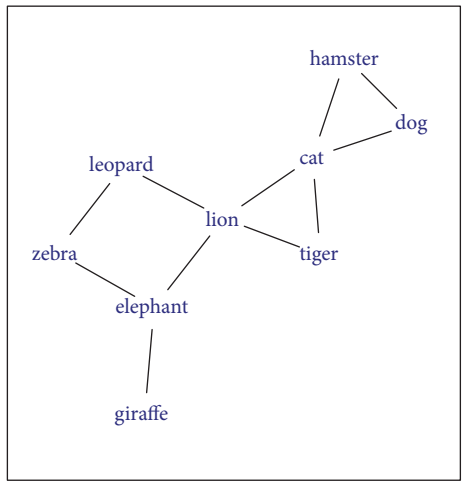
3.1. Cognitive Processes in Networks. A key advantage of the network science approach relevant for cognitive science is that it is possible to formalize processes that operate on the structure of a network. One of the earliest attempts to conceptualize cognitive processes on a network of nodes and their relationships focused on human semantic memory to explain an intricate taxonomy of human reasoning [140, 141]. When asked to verify statements such as *a robin is an animal* or *a robin is a bird*, participants typically take longer to verify the former statement as compared to the latter [142]. To account for this response time discrepancy, human semantic memory was assumed to exhibit a tree-like network organization, in which nodes represented concepts and edges represented whether words were elements of a higher order concept. In such a network, the triad, *robin*, *bird*, and *animal*, forms a line with *robin* being connected to *bird* and *bird* being connected *to animal*. Any reasoning process seeking to identify whether two concepts are, at least, indirectly connected via "is-contained-in" relationships would consequently have to traverse two edges to verify the first statement (i.e., the edges depicting *a robin is a bird* and *birds are animals*) but only one edge to verify the second (i.e., the edge that depicts *a robin is a bird*). This model of network processing represents one of the classic examples explaining behavioral phenomena as a combination of an underlying network representation and a process operating on the network.

Later, Collins and Loftus [27] generalized the insight that network structure in tandem with a retrieval process can account for human behavior and developed the theory of spreading activation in semantic memory (see also [143]). The key contribution of this theory was to make explicit the process operating on a semantic network when executing a cognitive task and to link the outputs of that process to participant performance and response times. Specifically, in the spreading activation theory it was assumed that reading or thinking about a concept would activate the concept and that this activation would spread to neighboring concepts in the network, priming related concepts and making them easier to retrieve. The process of spreading activation proposed by Collins and Loftus [27] implicitly assumes the presence of some cognitive resource (i.e., activation) that can be assigned to specific nodes, spread among connected nodes in a predefined network, and decay over time (as formally implemented in [143, 144]; see also [137]). This spread of activation quickly decays over time and distance in the semantic network [145]. Overall, the success of the spreading activation account demonstrates the significance of formalizing a process that captures search within a cognitive representation.

Network science provides new ways of expanding the original conceptualization of spreading activation by Collins and Loftus [27] by formalizing diffusion models over network representations. Diffusion processes on a network have been independently used to extensively study and predict epidemics of disease spread and of contagious ideas in a population (e.g., [146, 147]). The models of diffusion developed in these domains can be similarly used and adapted to study how information or activation "spreads" in a cognitive network. Although different implementations of network diffusion models exist, one core idea that illustrates the power of such network process models is the notion of a random walk on a network. A random walk model is a naïve search process that moves from node to node as function of a set of transition probabilities specifying the probability of moving from one node to any of its directly connected neighboring nodes. Note that when combined with a decay parameter, a random walk model becomes similar to a model of spreading activation [6]. However, despite their similarity, there are noteworthy differences in the implementations and goals of spreading activation in the tradition of Collin and Loftus [27] and random walk models by network scientists. Specifically, random walk models produce individual paths taken by the walk (e.g., an ordered list of words; see Figure 4), whereas a process of spreading activation produces a pattern of activation levels among nodes in the network and how they change over time. In contrast to random walks, spreading activation represents the aggregate, long-run behavior arising from an underlying basic process which could be modeled as a random walk. Diffusion models provide a means of exploring existing cognitive theories and insights such as spreading activation and have the ability to account for individual decision processes as well as describe the functioning of various cognitive systems. In recent years, various empirical studies have demonstrated how memory search can be modeled as a random walk process over semantic memory and have shown predictive power in accounting for human behavior [33, 148–151].

In the rest of this section, we showcase in greater detail recent empirical work showing how behavioral data from experiments can be used to provide a deeper understanding of the interaction of structure and processes that occur in cognitive and lexical networks. Each subsequent section focuses on a different cognitive domain—specifically, lexical retrieval, creative processes, and search and navigation in cognitive networks—that draw on either the process of random walks or the theoretical construct of spreading activation to account for relevant behavioral findings. The final section focuses on a key debate in the cognitive science regarding the complexities of disentangling structure and process in the domain of cognitive search and attempts to show how network science methods can contribute to this important debate.

3.2. Lexical Retrieval. The theory of spreading activation as proposed by Collins and Loftus [27] offers a blueprint for a mechanistic explanation to several psycholinguistic studies examining similarity effects on language processing. As mentioned in Section 2.2.1 (Microscopic Network Measures),



(a) Semantic network of animal category.

Censored random walk:

List 1: dog, cat, lion, elephant, ~~lion~~, tiger, ~~lion~~...
List 2: cat, hamster, dog, cat, lion, leopard, ~~lion~~...
List 3: elephant, zebra, leopard, lion, ~~leopard~~...

(b) Paths of censored random walks on a semantic network. Nodes that were visited more than once are "censored"

FIGURE 4: An example of a random walk process on a semantic network to account for responses in a fluency task. Adapted from Zemla & Austerweil [33].

network measures such as degree, local clustering coefficient, closeness centrality, and many others can be calculated for individual words. Critically these features may determine how activation spreads throughout a network and, thus, influences behavioral performance in psycholinguistic and memory tasks. Studies that have investigated such behavioral tasks have found that words with more clustered neighborhoods (words with high c_i values) are more slowly responded to in spoken [30] and visual word recognition tasks [48, 92], and more slowly produced [93] than words in less clustered neighborhoods. In these psycholinguistic tasks, participants typically are presented with, and respond to, words that have either high or low values on a particular network measure (e.g., higher or lower clustering coefficients, or higher or lower closeness centralities) in order to detect its influence on the efficiency of lexical retrieval that is typically indicated by faster reaction times and higher accuracy. Examples of such psycholinguistic tasks include lexical decision, where participants decide whether the presented stimulus represented a real word or a nonsense word, and speeded naming, where participants read out loud the presented word. Fast reaction times and high accuracy rates on these tasks indicate greater efficiency of lexical retrieval (i.e., a processing benefit or advantage). To account for clustering coefficient effects in word recognition of words versus nonwords, Chan and Vitevitch [30] provide a theoretical account that assumes a spreading activation process operating on a phonological network, where a word's structural characteristics affect how activation spreads through the network. Words with low clustering coefficients are hypothesized to receive more activation (and hence a processing advantage) from its neighbors as compared to those with high clustering coefficients because activation in the latter case is more likely to be shared among neighbors resulting in lower levels of activation for the target word. This account of a possible mechanism for word confusability within a theoretical spreading activation framework was further formalized and validated in computer simulations conducted by Vitevitch, Ercal and Adagarla ([152]; see also [137]).

Although spreading activation process as discussed above is mainly related to the local structural properties of words, the spreading activation process can also be used to account for findings showing that global structural aspects of words in the network influence lexical retrieval. Such findings suggest that closeness centrality of words influence spoken and visual word recognition [48, 97], that the size of the component that words reside in (i.e., whether the target word is in the largest connected component or an isolate) influences spoken word recognition, serial recall, and picture naming [106, 107], and that assortative mixing by degree, the tendency for nodes with similar degrees to be connected to each other, influences failures in lexical retrieval [103]. Among these results, the finding of a processing advantage for high closeness centrality words in spoken word recognition [97] is especially interesting as this finding goes against the general finding that greater similarity does not tend to help recognition (e.g., degree and local clustering effects result in poorer recognition; [30, 89]). To account for the processing advantage for high closeness centrality words, however, we can again draw on spreading activation theory. Goldstein and Vitevitch [97] suggest that, as a result of the activation of lexical representations spreading over time, high closeness centrality words will accumulate more activation than less central words due to their topologically advantageous location in the network. This higher accumulation of long-term activation provides an explanation for the processing benefit of words with high closeness centrality that counteract the more immediate, negative effects of neighborhood size and clustering—namely, that words that reside in structurally important locations in the network are suggested to have higher resting activation levels due to residual activation from other words in the network.

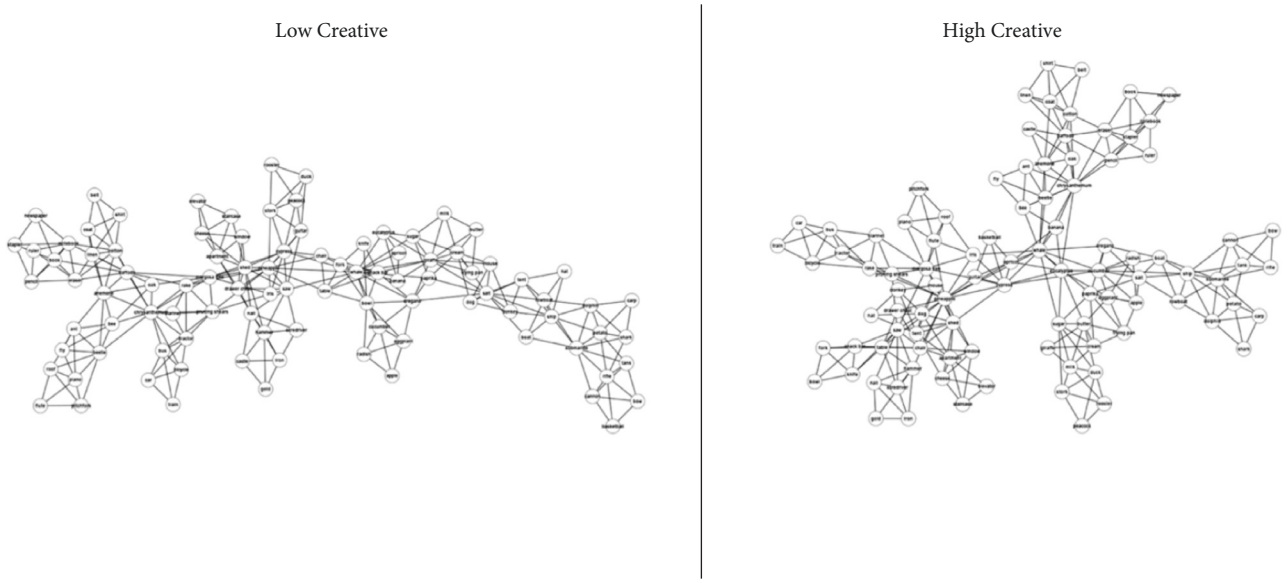


Figure 5: 2-D representation of a 96 node (cue words) semantic network of individuals with high and low levels of creative ability. Edges represent a binary, symmetric relation between nodes. Adapted from Kenett et al. [34].

3.3. Creative Processes. Another cognitive domain where spreading activation has proven to be a powerful framework is creativity. Theories of the creative process attribute a key role to spreading activation over memory to account for individual differences in creativity [153–155]. Such theories propose that creativity is related to the ability to combine together weakly related (i.e., distant) concepts into new concepts that are both novel and appropriate. On average, the further away the combined components are in a semantic network, the more novel the new concept will be [154]. A number of studies have applied network science methodologies to examine the idea that creativity involves the combination of distant concepts. These studies have shown how differences in semantic memory structure relate to individual differences in creativity, both at the group level (Figure 5; [34]) and at the individual level [121]. In group-level studies, semantic networks of low and high creative individuals were measured using a continuous free association task, which required participants to generate as many associations they can think of in one minute to a list of cue words. These responses were then used to determine the overlap of associations for all pairs of cue words (cf. [27]), which gave rise to separate semantic networks of low and high creative individuals [34, 40]. Individual-level networks were measured directly using relatedness judgments for all possible pairs of a list of target words. This approach allowed linking the resulting semantic networks to individual-level measures of creative ability [121].

Focusing on the macro- and mesolevel structure, these studies found that higher levels of creativity were associated with semantic networks that exhibit higher clustering coefficients, smaller average shortest path lengths, and lower modularity in community structure. The authors argue that the structural macroscopic properties of the semantic network of higher creative individuals facilitate the spread of activation across network communities in their semantic memory, leading to the generation of more novel ideas as activation reaches nodes in the network that are further apart from each other. Indeed, simulated random walks over the semantic networks of individuals of low and high creative ability revealed that the simulated search processes over the semantic network of highly creative individuals reached distant nodes and nodes with weaker relations ([156]; see [157], for further empirical support). These findings demonstrate how network science can quantitatively investigate how the structure of a cognitive system constrains processing on the representation [153, 158].

3.4. Mental Navigation and Cognitive Search. A large body of research has shown that people search their internal cognitive spaces in similar ways as they would in an external, physical space (e.g., [159–161]), implying that spatial and cognitive search processes have similar evolutionary roots [162]. Representing semantic memory as networks essentially produces a map of semantic space that allows one to mathematically trace search processes as paths in a network and make predictions about how structural properties of such network maps influence cognitive search behavior.

In one study, Iyengar et al. [99] had participants play a word-morph game that required them to transform one word into another by only changing one letter at the time (e.g., *ball- -tall- -tale- -take*). This word game is analogous to finding a path in a lexical network of word forms. The results showed that over time participants began to actively utilize network landmarks, hub words, and words of high closeness centrality, to drastically improve their performance both in terms of time and minimal number of word forms used. In another study, individuals were able to successfully identify the shortest path between a start word to a target word within a predefined semantic network based on free association norms [150, 163]. This result indicates that people

are able to actively exploit both global and local structure of semantic memory in order to estimate the distance between two words and to use that information to guide their search (see also, [164–166]).

Similar conclusions were drawn from more open-ended search tasks requiring individuals to actively explore their mental representations, such as category and letter fluency tasks. In verbal fluency tasks, individuals are required to retrieve from memory, in a fixed amount of time, as many elements belonging to a semantic category (e.g., name all the animals that you can think of) or words beginning with a specific letter (e.g., name all the words beginning with the letter 'S') as they can [167, 168]. Previous work has used behavioral data obtained from the fluency task to infer the structure of semantic networks via various estimation or inferential techniques (e.g., [33, 128, 169]) and to analyze and compare macro- and/or mesolevel network structure. This has led to important insights into the structural differences in semantic networks between younger and older adults [24, 170], less creative and highly creative individuals [118, 171], the first and second languages of bilinguals [172], and individuals with low or high openness to experience, a personality trait related to intellectual curiosity and an active imagination [173].

The sequences that individuals produce in verbal fluency tasks (in particular when listing items from a semantic category) can, however, also be used to study the search processes involved when individuals forage their mental representations. The finding that items with high semantic relatedness and many shared features tend to cluster together (e.g., [174]) was proposed to be related to an active search process that dynamically switches between retrieval cues [160, 161, 175, 176]. Other work has shown that micro (node-level) information such as PageRank centrality was most predictive of word recall [41] and that process models such as random walks could account for verbal fluency data ([33, 148]; Goñi et al., 2010). A prevailing question in this line of research is whether such search processes are better accounted for by a foraging process or a random walk remains open to debate and research ([177]; see next section).

Collectively, the studies discussed above provide important insights into how people navigate and retrieve information from their semantic and linguistic networks. Strong behavioral evidence and model-based approaches show that people are able to successfully search their mental representations in flexible ways to accomplish a variety of cognitive tasks, such as converting a word to another word, searching the semantic space as quickly and efficiently as possible, or generating creative ideas. The ability of network processes to capture behavioral differences suggests that the network science perspective provides ways to formulate testable hypotheses with regard to how individuals are accessing and navigating these cognitive structures.

3.5. Disentangling Structure and Process. A critical and open debate in cognitive network science is whether insights into the network structure and representations can be obtained independently from the retrieval processes operating upon it. This debate arises from the fact that both the underlying structure and the process operating on the representation are flexible enough to produce a wide array of behavior. Consider, for example, a verbal fluency task on the country category (i.e., name as many countries as you can think of), where participants can use various retrieval cues to help them complete the task. For a participant who uses the geographic relationships between countries as a retrieval cue, *France* and *Spain* would be very close, whereas *France* and *French Polynesia* would be very distant. On the other hand, for a participant who uses the phonological (sound-based) relationships among country names as a retrieval cue, *France* and *Spain* are now distant, whereas *France* and *French Polynesia* would be close as they share the same first sound. Research has found that individuals can flexibly switch between such retrieval strategies, essentially creating "wormholes" in memory, where shortcuts are (momentarily) created in the memory space to connect previously distant concepts ([161]; see also, [178]).

The indeterminacy problem, which is associated with the notion that both representation and retrieval processes can be powerful explanations of human behavior, has been the focus of a recent debate on models of the verbal fluency task. In modeling verbal fluency using a semantic space extracted from a text corpus, Hills et al. found evidence for an active search process that dynamically switches between subcategories of the semantic space [160, 179]. Shortly after, Abbott, Austerweil, and Griffiths [148] argued that a simple random walk model operating on a semantic network constructed from free associations—a model that does *not* require a switching process—is equally plausible as a mechanism of search in verbal fluency tasks, suggesting that a simpler model could reproduce the original results found by Hills, Jones, and Todd [160]. However, Abbott et al. evaluated the verbal fluency search on a network estimated from free association data, which—as subsequently argued by Jones, Hills, and Todd [180]—may already contain traces of the underlying search processes in semantic memory, rendering it unnecessary to account for such traces using an elaborate search model.

There have been a few attempts to disentangle process and structure (e.g., [177, 181]). One approach to address this issue is by developing research designs that directly compare representation-based hypotheses against process-based hypotheses on the same type of behavioral data. One such attempt was recently conducted by Kenett et al. [171], who examined the relationships between semantic network structure, creative ability, and intelligence in a large sample of individuals. The sample of participants were divided according to two dimensions – low/high creativity and low/high intelligence, and the animal category networks of all groups were estimated and compared. The authors found that creative ability and intelligence were associated with different structural aspects of the semantic network as estimated from the verbal fluency task. Specifically, intelligence was related to higher average shortest path length and modularity, whereas creativity was related to higher "small worldness" properties of the semantic network. Taking intelligence as a proxy for the contribution of cognitive control processes such as attention and working memory, and creativity as an emergent

property of the underlying network structure, these results suggest different contributions for process and representations and demonstrate one possible path for disentangling the two.

The bottom line of this open debate of the influence of structure and process on cognitive representations is that unless one of the two is clearly identified, it is difficult, if not impossible, to make strong inferences from data about the other. Currently, a growing body of work focuses on examining the reliability and reproducibility of estimating networks in both cognitive and psychological networks [24, 33, 69, 170, 173] and this task of network estimation remains an important challenge for network science approaches to understand the details of a given cognitive system. Another area of avid development and a promising route to addressing the indeterminacy problem is the employment of plausible learning models to learn a semantic structure from digitized text and images rather than responses in behavioral tasks. Structure created in this way will contain little trace of the processes operating on it, and may allow for cleaner separation of the contribution of structure and process [180]. Moreover, many of these learning models, such as the BEAGLE model [182], process the environmental input incrementally, principally allowing for the modeling of developmental changes in the network structure.

3.6. Summary. In this section, we discussed recent empirical work that provided a deeper understanding of the interaction of structure and processes that occur in cognitive and lexical networks. We focused on the cognitive domains of lexical retrieval, cognitive search, and creativity to illustrate how the process of random walks or spreading activation can be implemented in a cognitive network representation to account for a variety of behavioral phenomena and offer novel insights and testable predictions of how individuals might be accomplishing these tasks. These studies not only emphasize the importance of considering how the structure of the underlying network interacts with processes operating in it, but also the complexities of disentangling structure and processes, particularly in the domain of cognitive search. We conclude that the consideration of structure and process emerges naturally from a network science approach by compelling researchers to explicitly define and model the relationship between structure and process in order to account for human behavior and cognition. Formalizing the relationship between process and structure enriches our theoretical understanding of the interplay between cognitive processes and cognitive structures in various domains.

4. Network Dynamics across Multiple Timescales

Conceptualizing cognitive systems in terms of a network representation not only motivates cognitive scientists to think more explicitly about the structure of cognitive systems (Section 2: Network Representations of Cognitive Systems) and how cognition might be captured by processes operating on a network (Section 3: Processes in Cognitive Networks), but also stimulates the question of how a particular structure arose and how this structure develops and changes across time.

In this section, we posit that semantic and lexical networks are inherently dynamic—the structure of such cognitive systems changes at multiple timescales and in response to (i) linguistic input and experiences that reflect the long-term accumulation of knowledge and (ii) exposure to experimental tasks and manipulations that trigger more immediate functional changes in semantic memory [26, 183]. The first part of this section focuses on developing semantic networks to capture the process of language acquisition. The second part focuses on semantic networks of older adults. Finally, the third part focuses on individual differences in semantic networks. Although it is possible to examine the network dynamics of the language system itself, i.e., how the structure of language has evolved over time (see [184, 185], for examples), we focus our discussion of network dynamics on cognitive network structures that are relevant for semantic memory and lexical access.

4.1. Developmental Networks. Rather than focusing on a single snapshot of a network in time, some recent investigations include a temporal dimension to quantify and elucidate how network representations of individuals change across the lifespan. For instance, Hills et al. [35, 186] used normative data from the MacArthur-Bates Communicative Development Inventory (CDI; [187]), a vocabulary checklist completed by parents to indicate the words produced by their child across development, to empirically study normative language development using networks. These data can be used to create semantic networks that develop over the course of language acquisition by placing edges between words that an average child knows at a given time point. Here edges can represent semantic associations that are based on adult free association norms (e.g., [31]) or cooccurrences in child-directed speech [188]. Semantic networks created in this way conceptualize a normative child's semantic knowledge at various time points during development (see [188–190], for modeling of individual children's semantic network acquisition). To capture development over time, Hills and colleagues performed a statistical comparison of three different models of network growth (see Figure 6): preferential attachment, preferential acquisition, and lure of the associates. In the spirit of Barabási & Albert's original model [134], the preferential attachment model predicts that words that connect to well-connected words already known by the (normative) child will be learned earlier. The preferential acquisition model predicts that words are learned earlier if they are highly connected in the learning environment as approximated by the full (adult) semantic network. Finally, the lure of associates model predicts that new words are more likely to be learned if they result in the addition of more connections to the network of words already known by the child.

Comparing these models, Hills et al. found that the preferential attachment model was in fact *not* a good fit to the normative CDI data and that the preferential acquisition model was able to best account for vocabulary growth in early semantic networks [35, 186, 191]. These modeling results highlight that the learning environment plays an important

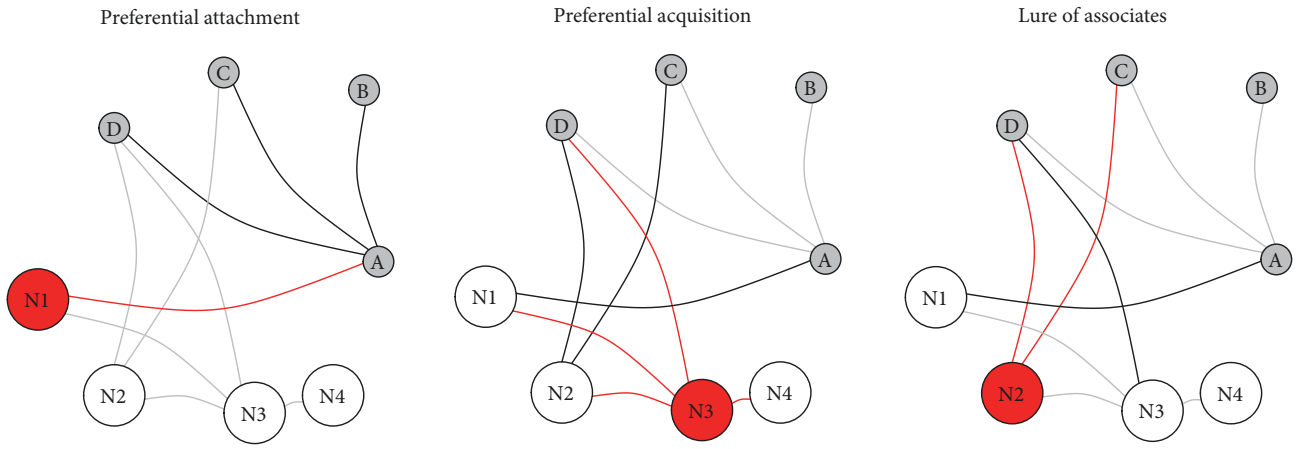


Figure 6: The three growth models of semantic networks. Note that the models make different predictions about which words are more likely to be acquired first despite having the same underlying network structure. Smaller, grey nodes indicate the words already known to the normative child and larger white nodes are words that are not yet learned. The red node is more likely to be acquired before the other nodes based on the model's prediction. Adapted from Hills et al. [35].

role in language acquisition—words that occur in many different contexts (and hence are well-connected) in the learning environment are more likely to be acquired before less connected words in the learning environment. Additionally, these results also demonstrate that growth models other than preferential attachment (and other related variations) can lead to scale-free degree distributions of semantic networks.

The growth mechanisms underlying semantic and language networks remain, however, imperfectly understood (for an overview, see [192]). For instance, Hills et al. [186] found that the preferential acquisition model accounted for growth of semantic networks with edges constructed from free associations, but not for semantic networks with edges constructed from shared features, suggesting that different growth processes may govern different aspects of language. In phonological networks, psycholinguistic evidence seems to support the lure of associates model [193, 194], although this model remains to be empirically tested against other growth models such as preferential attachment and preferential acquisition. The difference between semantic and phonological networks could provide unique insight into the acquisition process of children—for example, phonological information may be more constrained by phonological form characteristics of the child's current vocabulary, whereas semantic knowledge may be readily observed in the physical environment, accounting for the difference in the contribution of the child's current vocabulary in phonological and semantic domains (e.g., see [66]).

Studies have recently pursued new approaches by examining language acquisition in terms of feature networks [195, 196], or multiplex networks representing both semantic and phonological information [65, 66]. One particularly informative approach has been to study atypical acquisition processes. Using the previously described approach of converting vocabulary checklist data into semantic networks and then further analyzing its network structure, Beckage, Smith, and Hills [188] examined the semantic networks of children who were classified as late talkers and found that the semantic networks of late talkers had, on average, higher average path length and lower clustering as compared to the semantic networks of typically developing children, even after controlling for differences in network size or the age of the child. These differences between typically developing children and those with risk of language impairment suggested a maladaptive tendency in late talkers to acquire "odd" words that were less connected in the semantic network (see also [49]). Studying the structure of semantic networks of children with cochlear implants, Kenett et al. [169] also found differences in the semantic network of children with cochlear implants as compared to the semantic network of typically developing children. Specifically, they observed shorter average path lengths for children with cochlear implants relative to typically developing children, suggesting an underdevelopment of the semantic network due to impoverished input.

4.2. Aging Networks. The development of semantic networks does not halt with onset of adulthood (see [8], for a review). Language learning and change continues throughout the lifespan. Wulff and colleagues [24, 170] examined the semantic networks of younger and older adults (as estimated from semantic fluency data and similarity ratings obtained from both populations) and found that both the aggregate and individual older adults' networks exhibited lower entropy in the degree distribution, larger average shortest path lengths, and smaller clustering coefficients relative to the aggregate and individual networks of younger adults. Using free associations obtained from a cross-sectional sample across the lifespan to estimate semantic networks for groups of young, middle-aged, and older adults, Dubossarsky, De Deyne, and Hills [36] similarly found that structure of the semantic network in early life reverses, in parts, in later life (see Figure 7; for similar results see [197]). Behavioral research on cognitive aging usually finds that older adults take more time and perform worse on variety of cognitive tasks involving memory, concentration, and reasoning than younger adults, which is commonly attributed to cognitive slowing [198, 199].

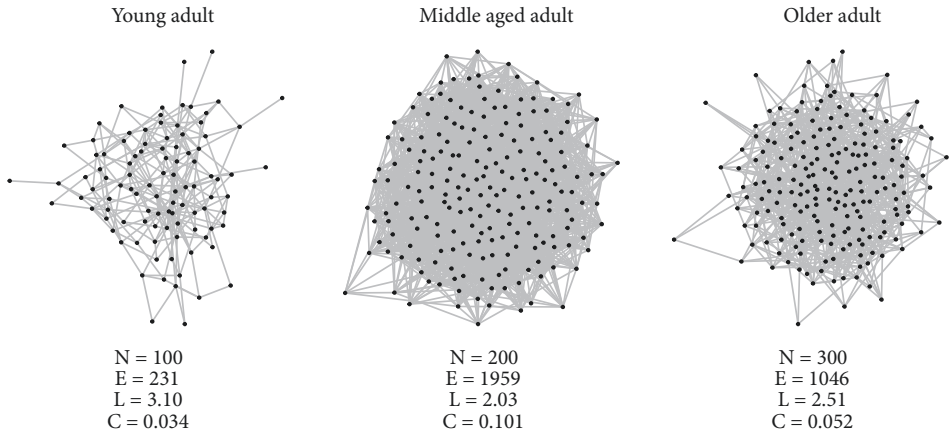


Figure 7: The structure of free association networks changes across the lifespan, with the youngest network on the left and the oldest network on the right. Over the lifespan, the network grows in size (i.e., the number of nodes, *N*, increases). The network is quite sparse in early life, becomes most densely connected early in adulthood, and becomes less connected in old age. Adapted from Dubossarsky, De Deyne, and Hills [36].

However, the findings of Dubossarsky et al. [36] and Wulff et al. [170] fuel an interesting hypothesis regarding age-related decline in cognitive function. Based on the observed differences in the structure of semantic networks, changes in the underlying representation of knowledge might be held partly accountable for behavioral changes in older adulthood [8, 23, 24, 200]. Older adults have access to a larger vocabulary, implying a larger network representation with more nodes and edges. Their larger network may exact higher search costs in accessing and navigating the representation which may account for the observed cognitive slowing. Assuming a process akin to spreading activation, it is further conceivable that other changes in structure affecting, for instance, the average path length or clustering coefficient, of older adults' networks may play an additional role. Thus, in the case of age-related cognitive decline, the application of a network approach has led to a promising rival explanation to established theories, paving the way for a new perspective on an important issue considering our aging society.

In light of the growing importance of age-related degenerative diseases, such as dementia and Alzheimer's disease, it is especially worthwhile to study models that may capture possible mechanisms underlying these changes in behavior and provide a link to memory representations. To understand the cognitive mechanisms underlying pathological decline, studies have focused on comparing healthy individuals to clinical patients. In one study, Borge-Holthoefer, Moreno, and Arenas [201] conducted simulations based on a network model of language degradation in order to account for hyperpriming among people with Alzheimer's disease. Hyperpriming refers to increased priming effects observed in patients with Alzheimer's as compared to healthy controls in naming and lexical decision tasks [202], despite the fact that Alzheimer's patients have more difficulties in other memory related tasks. Their simulations showed that a process of edge degradation provided a possible mechanism for why individuals with Alzheimer's disease show evidence of hyperpriming. Specifically, the effect appeared to be driven by the fact that weak (distinctive) associations were lost earlier than stronger (common) associations, leading to a loss of distinctiveness between primes and targets and an increase in priming effects for related words. This result suggests that changes in aging semantic networks are not simply the "inverse" of early development.

4.3. Individual Differences in Networks. The structural differences observed in semantic networks of younger and older adults, as discussed above, indicate that mental representations are shaped, to some extent, by the specific environmental input that individuals are exposed to [8]. This has broad implications beyond research on development and aging and specifically in the domain of personality and individual differences. If we accept the premise that the experiences of individuals can influence their mental representations, this naturally provides a plausible mechanistic explanation that could help us understand individual differences observed in the structure of mental representations—differences that reflect the accumulation of experiences over an individual's life. In line with this idea, research has demonstrated links between semantic network structure and personality. Recent work by Christensen, Kenett, Cotter, Beaty, and Silvia [173] found that the semantic networks of people with higher openness to experience (where semantic networks were inferred from verbal fluency responses and levels of openness were measured using an independently administered personality questionnaire) were more interconnected and better organized as compared to the semantic networks of people with lower openness to experience. One possible explanation for this result is that individuals who are open to experiences are drawn to, and subsequently exposed to, more diverse environments, which may encourage the formation of diverse associations and may thus account for the higher interconnectivity observed in their semantic networks.

The link between experience and the underlying network structure may also have implications for other domains, such as creativity and problem solving. As discussed earlier, higher creative people exhibit less modular semantic networks than lower creative people [34, 153], which is accompanied by a

reduction in path length and may facilitate insight problem solving [203]. These results suggest a codependent relationship between personality and mental representations, where individuals with different personalities seek out different experiences, which shapes their mental representations and, in turn, their behavior and personality.

To date, almost all cognitive network studies have estimated cognitive networks of groups. To truly examine individual differences in cognitive networks and how they may relate to other cognitive and psychological constructs, methods must be developed to estimate cognitive networks at the level of the individual. Only a few studies have attempted to estimate an individual's semantic network. Morais, Olsson, and Schooler [133] collected free associations from participants based on an associative "snow-ball" sampling approach, where participants provided associations to a set of "seed" cue words in the first iteration, and in the second iteration, associations to their own associative responses in the previous iteration, and so on. These data were used to estimate semantic networks of each individual. Austerweil and colleagues [33, 204] developed a computational approach to estimate individual semantic networks from semantic fluency data, based on a censored random walk model of memory retrieval (see also [24]).

Other studies estimated individual semantic networks based on semantic relatedness ratings [24, 121] and related the structural properties of these individual semantic networks to individual differences with respect to the individual's age ([24]; see Section 4.2), or the individual's creative ability and intelligence measures [121]. The latter study replicated the group-based network analysis of Kenett et al. [34] described above, by finding a positive correlation with CC, a negative correlation with ASPL, and a trending negative correlation with Q and creative abilities. Furthermore, the authors showed how the ASPL of each participant's network and their measure of fluid intelligence predicted creative thinking [121], providing further evidence demonstrating how process and structure of memory might be independently measured.

Although additional methodological development is required, these methods provide researchers with the means to estimate individual cognitive networks. Current theories of semantic memory view semantic memory as a dynamic system that is contingent on different contexts (e.g., environment, age) and individual differences (e.g., creativity, personality traits), with short- and long-term effects on semantic memory reflecting the cognitive processes operating on semantic representations [8, 26, 183]. Therefore, quantifying the properties of cognitive systems, such as semantic memory, at the individual level can greatly advance our understanding of their complex structure and dynamics.

4.4. Summary. The tools of network science have revealed interesting differences in the network structure of various populations (e.g., younger vs. older adults, healthy vs. clinical populations, individuals with low openness to experience vs. individuals with high openness to experience). Such findings support the idea that network representations and processes are inherently dynamic, compelling cognitive scientists to formalize theoretical and algorithmic explanations for how these differences emerge in the first place. By combining process or growth models provided by network science with empirical data from the cognitive sciences, researchers can formalize relationships between (network) representation and processes that operate within the representation and construct models that take into account the contributions of external factors, such as the linguistic environment of young language learners, transient changes that are due to creative abilities, or the accumulative effects of semantic knowledge acquired in a person's lifetime. The framework of process and structure can offer explanations and predictions of changes in the structure of cognitive representation, which in turn affects our understanding of language and cognitive processes in a continual, interacting cycle.

5. Summary and Conclusions

In this review, we demonstrate the usefulness of the network science approach to the study of cognition in at least three ways.

(1) Network Science Provides a Quantitative Approach to Represent Cognitive Systems. To demonstrate the quantitative power of network science in describing cognitive systems, Section 2 discussed how networks can represent a variety of cognitive systems, including language, semantic memory, personality traits, and the language environment of individuals (see Table 1). Furthermore, we highlighted a host of network measures that are available to the researcher when he or she commits to the theoretical decision of representing the cognitive system of interest as a network. We reviewed previous research using these tools to characterize the structure and behavior of networks on the micro-, meso-, and macroscopic levels in order to derive novel insights.

(2) Network Science Facilitates a Deeper Understanding of Human Cognition by Allowing Researchers to Consider How Network Structure and the Processes Operating on the Network Structure Interact to Produce Behavioral Phenomena. Section 3 focused on processes operating on networks. Network representations of cognitive systems, particularly in the area of language and semantic memory, are often used to represent a latent mental structure that requires the assumption of some process to link the network to observable behavior. Adopting a network science approach naturally compels researchers to consider the interaction between structure and process, which has been especially useful in the three research domains discussed in Section 3 (lexical retrieval, creativity, and cognitive search). Finally, we briefly discuss the difficulties in dissociating structure and process, particularly as it relates to the modeling of behavioral outputs in retrieval tasks from semantic memory and suggest ways in which network science methods can enrich the investigation of such cognitive phenomena.

(3) Network Science Provides a Framework to Model Structural Changes in Cognitive Systems at Multiple Timescales. In Section 4, we discussed how network science can be used to study the development of cognitive systems, enabling a better

understanding of cognition at both the early and late stages of human life, as well as structural changes that occur at more immediate timescales, as related to higher order cognitive processes such as creative insight and problem solving. Such cognitive systems can be modeled as a dynamic network representation that changes in response to the accumulation of experiences and linguistic input. The research discussed in this section demonstrates how network science approaches can be used to quantify structural changes and the dynamics of cognitive systems across different timescales.

5.1. Future Directions. It is clear from this review that network science approaches have contributed much to the study of human cognition. Indeed, this is reflected in the recent growth in methodological and computational tools designed specifically for the cognitive scientist to model, analyze, and visualize cognitive and language networks (see Section 2.3: *Methodological Tools and Resources for Cognitive Network Analysis*, for a compilation of methodological tools commonly used in the field of cognitive network science and by many of the studies described in this review). But it is important to emphasize that cognitive network science is a relatively young field and many methodological and theoretical challenges remain to be addressed. For example, as previously discussed in Section 3.5, to what extent can specific aspects of the network structure and the processes operating in the network be disentangled? Below we briefly highlight three milestones that cognitive network science needs to achieve in order to become a mature research paradigm in the cognitive sciences and in the network science community.

One critical milestone needed in cognitive network science is the development of inferential methodologies to analyze empirical networks—that is, networks that are inferred from behavioral data. Currently, statistical models that allow for hypothesis testing when comparing empirical networks remain a major challenge. This challenge is mainly due to difficulties in estimating or collecting a large sample of empirical networks and the lack of accessible statistical methods to test differences in observed networks [205]. In these cases, bootstrapping methods over comparable networks might be a solution [206]. Similarly, issues in how to minimize spurious connections in psychological and cognitive networks are currently debated and require further methodological development [69].

Another crucial milestone is the development of methods to represent psychological and cognitive networks at the individual level (see Section 4.3). Psychological and cognitive constructs vary across individuals and aggregating across participants in group-based cognitive network analysis may conceal nuanced differences across individuals. While there has been some attempts at representing individual semantic networks [24, 33, 121, 133, 204], developing a reliable and easy-to-apply methodology to represent an individual's semantic network will allow researchers to design studies that relate individual representations to other cognitive and neural measures.

Finally, a third milestone is the development of new network science methodologies to quantitatively study specific theoretical issues across different cognitive domains. Two such examples were briefly discussed in the review: the application of multiplex network analysis to examine how different cognitive domains interact (e.g., [65, 66, 207]) and the application of percolation theory to study cognitive phenomena such as memory decline or flexibility of thought (e.g., [119, 201]). Cognitive network science could also benefit from adopting and contributing to state-of-the-art network methodologies used to study neural systems and brain dynamics, such as network dynamic analysis [208] and network control theory [209–211]. Network dynamic analysis examines time varying community assignments over brain functional connectivity networks and has attributed state flexibility—variation of assignment of a brain region to a specific community across time—to capacities such as motor-skill learning and language comprehension [208]. Network control theory quantifies the extent that different nodes in a network drive the dynamics over the network. Recent studies have applied network control theory to the analysis of white-matter connectivity networks to examine the roles of different brain regions in driving neural dynamics [209–211]. Importing such state-of-the-art methods to the cognitive domain could greatly advance the study of dynamics in cognitive networks.

5.2. Cognitive Network Science: A New Frontier. The aim of this review was to highlight and emphasize the feasibility and significance of applying network science methodologies to study cognition. Such applications allow the quantification of theoretical cognitive constructs and direct examination of cognitive theories in domains such as language and memory. The cognitive network framework also provides a means to model and formalize theoretical and mathematical descriptions of dynamics operating in cognitive systems. The work conducted in this new field of cognitive network science has already provided many novel insights on cognitive issues such as lexical retrieval, language acquisition, memory search and retrieval, bilingualism, learning, creativity, personality traits, and clinical populations.

In this review, we focused on lexical and semantic networks that capture various types of relationships between words. However, as alluded to where relevant research was available, the usefulness of network science in cognitive science is by no means limited to the domains of language and memory. Network science is a general-purpose toolbox that can be used to study many types of cognitive systems provided that it is indeed theoretically meaningful for these systems to be represented as a network. This further implies that the usefulness of network science depends on both the problem at hand and, of course, the researcher, who will make theoretically motivated decisions. These decisions include, specifically, (i) what aspects of the problem can and should be represented as a network, (ii) what tools and measures should be applied to analyze the network representation, and (iii) how to link network structure and process to offer meaningful insights into empirical and behavioral data. Network science is no panacea for the challenges faced by cognitive research, but when used appropriately, network science can produce important and novel insights for cognitive research, as demonstrated by the vast array

of research on cognitive network science covered in this review.

Our understanding of any cognitive or language-related process is necessarily incomplete if we do not consider the structural properties of the cognitive system that the process is occurring in. Network science provides cognitive scientists with a well-studied and formal language to quantify and study the structure of these cognitive systems. On the other hand, the cognitive science community has developed a suite of experimental tasks that can provide crucial behavioral evidence that constrains and informs network models of cognition. The judicious combination of these two approaches will lead to continued insights into a variety of behavioral and cognitive phenomena and strengthen psychological theories of lexical access, memory retrieval, cognitive search, language acquisition, cognitive decline, and creativity, as well as many other related domains of cognitive science that will benefit from the application of network science methods.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Cynthia S. Q. Siew is supported by the Overseas Postdoctoral Fellowship from the National University of Singapore.

References

[1] A.-L. Barabási, "Network science: Luck or reason," *Nature*, vol. 489, no. 7417, pp. 507-508, 2012.

[2] A.-L. Barabási, *Network Science*, Cambridge University Press, Cambridge, UK, 2016.

[3] O. Sporns, *Networks of The Brain*, M.I.T. Press, Cambridge, UK, 2011.

[4] A. Baronchelli, R. Ferrer-i-Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen, "Networks in Cognitive Science," *Trends in Cognitive Sciences*, vol. 17, no. 7, pp. 348–360, 2013.

[5] N. M. Beckage and E. Colunga, "Language networks as models of cognition: Understanding cognition through language," in *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, A. Mehler, P. Blanchard, and B. Job, Eds., pp. 3–28, Springer, 2015.

[6] S. De Deyne, Y. N. Kenett, D. Anaki, M. Faust, and D. J. Navarro, "Large-scale network representations of semantics in the mental lexicon," in *Big Data in Cognitive Science: from Methods to Insights*, N. M. Jones, Ed., pp. 174–202, Psychology Press: Taylor Francis, New York, NY, USA, 2016.

[7] R. V. Solé, B. Corominas-Murtra, S. Valverde, and L. Steels, "Language networks: their structure, function, and evolution," *Complexity*, vol. 15, no. 6, pp. 20–26, 2010.

[8] D. U. Wulff, S. De Deyne, M. N. Jones, and R. Mata, "The Aging Lexicon Consortium. New perspectives on the aging lexicon," *Trends in Cognitive Science*, 2019.

[9] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. W. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[10] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neuroscience*, vol. 20, no. 3, pp. 353–364, 2017.

[11] J. Borge-Holthoefer and A. Arenas, "Semantic networks: Structure and dynamics," *Entropy*, vol. 12, no. 5, pp. 1264–1302, 2010.

[12] M. N. Jones, J. Willits, and S. Dennis, "Models of semantic memory," in *Oxford Handbook of Mathematical and Computational Psychology*, J. Busemeyer and J. Townsend, Eds., pp. 232–254, Oxford University Press, Oxford, UK, 2015.

[13] J. Aitchison, *Words in the Mind: An Introduction to The Mental Lexicon*, John Wiley & Sons, 4th edition, 2012.

[14] K. Forster, "Memory-addressing Mechanisms and Lexical Access," *Advances in Psychology*, vol. 94, no. C, pp. 413–434, 1992.

[15] J. R. Anderson, "A Simple Theory of Complex Cognition," *American Psychologist*, vol. 51, no. 4, pp. 355–365, 1996.

[16] G. S. Dell, F. Chang, and Z. M. Griffin, "Connectionist models of language production: Lexical access and grammatical encoding," *Cognitive Science*, vol. 23, no. 4, pp. 517–542, 1999.

[17] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory," *Psychological Review*, vol. 102, no. 3, pp. 419–457, 1995.

[18] M. S. Seidenberg and J. L. McClelland, "A Distributed, Developmental Model of Word Recognition and Naming," *Psychological Review*, vol. 96, no. 4, pp. 523–568, 1989.

[19] N. Chater and C. D. Manning, "Probabilistic models of language processing and acquisition," *Trends in Cognitive Sciences*, vol. 10, no. 7, pp. 335–344, 2006.

[20] T. L. Griffiths, N. Chater, C. Kemp, A. Perfors, and J. B. Tenenbaum, "Probabilistic models of cognition: exploring representations and inductive biases," *Trends in Cognitive Sciences*, vol. 14, no. 8, pp. 357–364, 2010.

[21] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and Brain Sciences*, vol. 40, 2017.

[22] B. Hart and T. R. Risley, "Meaningful differences in the everyday experience of young American children," *Paul H Brookes Publishing*, 1995.

[23] M. Ramscar, P. Hendrix, C. Shaoul, P. Milin, and R. H. Baayen, "The myth of cognitive decline: Non-linear dynamics of lifelong learning," *Topics in Cognitive Science*, vol. 6, no. 1, pp. 5–42, 2014.

[24] D. U. Wulff, T. Hills, and R. Mata, "Structural differences in the semantic networks of younger and older adults," *PsyArXiv*, 2018.

[25] P. B. Baltes and U. Lindenberger, "Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging?" *Psychology and Aging*, vol. 12, no. 1, pp. 12–21, 1997.

[26] E. Yee and S. L. Thompson-Schill, "Putting concepts into context," *Psychonomic Bulletin & Review*, vol. 23, no. 4, pp. 1015–1027, 2016.

[27] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing," *Psychological Review*, vol. 82, no. 6, pp. 407–428, 1975.

[28] M. Steyvers and J. B. Tenenbaum, "The large-scale structure of semantic networks: statistical analyses and a model of semantic growth," *Cognitive Science*, vol. 29, no. 1, pp. 41–78, 2005.

[29] D. Lazer, A. S. Pentland, L. Adamic et al., "Life in the network: The coming age of computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[30] K. Y. Chan and M. S. Vitevitch, "The influence of the phonological neighborhood clustering coefficient on spoken word recognition," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 35, no. 6, pp. 1934–1949, 2009.

[31] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber, "The University of South Florida free association, rhyme, and word fragment norms," *Behavior Research Methods, Instruments, and Computers*, vol. 36, no. 3, pp. 402–407, 2004.

[32] M. S. Vitevitch, "What can graph theory tell us about word learning and lexical retrieval?" *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 2, pp. 408–422, 2008.

[33] J. C. Zemla and J. L. Austerweil, "Estimating semantic networks of groups and individuals from fluency data," *Computational Brain Behavior*, vol. 1, no. 1, pp. 36–58, 2018.

[34] Y. N. Kenett, D. Anaki, and M. Faust, "Investigating the structure of semantic networks in low and high creative persons," *Frontiers in Human Neuroscience*, vol. 8, pp. 1–16, 2014.

[35] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Categorical structure among shared features in networks of early-learned nouns," *Cognition*, vol. 112, no. 3, pp. 381–396, 2009.

[36] H. Dubossarsky, S. De Deyne, and T. T. Hills, "Quantifying the structure of free association networks across the lifespan," *Developmental Psychology*, vol. 53, no. 8, pp. 1560–1570, 2017.

[37] S. Pinker and R. Jackendoff, "The faculty of language: what's special about it?" *Cognition*, vol. 95, no. 2, pp. 201–236, 2005.

[38] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behavior Research Methods*, vol. 45, no. 4, pp. 1191–1207, 2013.

[39] S. De Deyne and G. Storms, "Word associations: Network and semantic properties," *Behavior Research Methods*, vol. 40, no. 1, pp. 213–231, 2008.

[40] Y. N. Kenett, D. Y. Kenett, E. Ben-Jacob, and M. Faust, "Global and local features of semantic networks: Evidence from the Hebrew mental lexicon," *PLoS ONE*, vol. 6, no. 8, Article ID e23912, 2011.

[41] T. L. Griffiths, M. Steyvers, and A. Firl, "Google and the mind: predicting fluency with PageRank," *Psychological Science*, vol. 18, no. 12, pp. 1069–1076, 2007.

[42] K. McRae, G. S. Cree, M. S. Seidenberg, and C. McNorgan, "Semantic feature production norms for a large set of living and nonliving things," *Behavior Research Methods*, vol. 37, no. 4, pp. 547–559, 2005.

[43] S. H. Solomon, J. D. Medaglia, and S. L. Thompson-Schill, "Implementing a concept network model," *Behavior Research Methods*.

[44] G. A. Miller, "WordNet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[45] S. Arbesman, S. H. Strogatz, and M. S. Vitevitch, "The structure of phonological networks across multiple languages," *International Journal of Bifurcation and Chaos*, vol. 20, no. 3, pp. 679–685, 2010.

[46] C. S. Q. Siew, "Community structure in the phonological network," *Frontiers in Psychology*, vol. 4, Article 553, 2013.

[47] C. T. Kello and B. C. Beltz, "Scale-free networks in phonological and orthographic wordform lexicons," in *Approaches to Phonological Complexity*, I. Chitoran, C. Coupe, E. Marsico, and F. Pellegrino, Eds., pp. 171–190, Mouton de Gruyter, Berlin, Germany, 2009.

[48] C. S. Siew, "The orthographic similarity structure of English words: Insights from network science," *Applied Network Science*, vol. 3, no. 1, article 13, 2018.

[49] J. Ke and Y. Yao, "Analysing language development from a network approach," *Journal of Quantitative Linguistics*, vol. 15, no. 1, pp. 70–99, 2008.

[50] R. F. I. Cancho and R. V. Solé, "The small world of human language," *Proceedings of the Royal Society B Biological Science*, vol. 268, no. 1482, pp. 2261–2265, 2001.

[51] O. Abramov and A. Mehler, "Automatic language classification by means of syntactic dependency networks," *Journal of Quantitative Linguistics*, vol. 18, no. 4, pp. 291–336, 2011.

[52] R. F. I. Cancho, R. R. V. Solé, and R. Köhler, "Patterns in syntactic dependency networks," *Physical Review E*, vol. 69, no. 5, Article ID 051915, 2004.

[53] J. Cong and H. Liu, "Approaching human language with complex networks," *Physics of Life Reviews*, vol. 11, no. 4, pp. 598–618, 2014.

[54] M. Choudhury, N. Ganguly, A. Maiti et al., "Modeling discrete combinatorial systems as alphabetic bipartite networks: Theory and applications," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 81, no. 3, Article ID 036103, 2010.

[55] A. Mukherjee, M. Choudhury, A. Basu, and N. Ganguly, "Modeling the co-occurrence principles of the consonant inventories: A complex network approach," *International Journal of Modern Physics C*, vol. 18, no. 2, pp. 281–295, 2007.

[56] S. Majerus, M. Van der Linden, L. Mulder, T. Meulemans, and F. Peters, "Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm," *Journal of Memory and Language*, vol. 51, no. 2, pp. 297–306, 2004.

[57] H. Small, "Visualizing science by citation mapping," *Journal of the Association for Information Science and Technology*, vol. 50, no. 9, pp. 799–813, 1999.

[58] H. T. Liu, "Statistical properties of Chinese semantic networks," *Chinese Science Bulletin*, vol. 54, no. 16, pp. 2781–2785, 2009.

[59] H. Liu, "The complexity of Chinese syntactic dependency networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 12, pp. 3048–3058, 2008.

[60] H. Liu, Y. Zhao, and W. Li, "Chinese syntactic and typological properties based on dependency syntactic treebanks," *Poznan Studies in Contemporary Linguistics*, vol. 45, no. 4, pp. 495–509, 2009.

[61] H. Liu and W. Li, "Language clusters based on linguistic complex networks," *Chinese Science Bulletin*, vol. 55, no. 30, pp. 3458–3465, 2011.

[62] H. Liu and J. Cong, "Language clustering with word co-occurrence networks based on parallel texts," *Chinese Science Bulletin*, vol. 58, no. 10, pp. 1139–1144, 2013.

[63] H. Liu and C. Xu, "Can syntactic networks indicate morphological complexity of a language?" *Europhysics Letters*, vol. 93, no. 2, p. 28005, 2011.

[64] S. Yu, H. Liu, and C. Xu, "Statistical properties of Chinese phonemic networks," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 7, pp. 1370–1380, 2011.

[65] M. Stella, N. M. Beckage, and M. Brede, "Multiplex lexical networks reveal patterns in early word acquisition in children," *Scientific Reports*, vol. 7, Article ID 46730, 2017.

[66] M. Stella, N. M. Beckage, M. Brede, and M. De Domenico, "Multiplex model of mental lexicon reveals explosive learning in humans," *Scientific Reports*, vol. 8, no. 1, Article ID 2259, 2018.

[67] D. Borsboom and A. O. J. Cramer, "Network analysis: An integrative approach to the structure of psychopathology," *Annual Review of Clinical Psychology*, vol. 9, pp. 91–121, 2013.

[68] E. I. Fried, C. D. van Borkulo, A. O. J. Cramer, L. Boschloo, R. A. Schoevers, and D. Borsboom, "Mental disorders as networks of problems: a review of recent insights," *Social Psychiatry and Psychiatric Epidemiology*, vol. 52, no. 1, pp. 1–10, 2017.

[69] A. P. Christensen, Y. N. Kenett, T. Aste, P. J. Silvia, and T. R. Kwapil, "Network structure of the wisconsin schizotypy scales–short forms: examining psychometric network filtering approaches," *Behavior Research Methods*, vol. 50, no. 6, pp. 2531–2550, 2018.

[70] K. T. Forbush, C. S. Q. Siew, and M. S. Vitevitch, "Application of network analysis to identify interactive systems of eating disorder psychopathology," *Psychological Medicine*, vol. 46, no. 12, pp. 2667–2677, 2016.

[71] R. J. McNally, D. J. Robinaugh, G. W. Y. Wu, L. Wang, M. K. Deserno, and D. Borsboom, "Mental disorders as causal systems: A network approach to posttraumatic stress disorder," *Clinical Psychological Science*, vol. 3, no. 6, pp. 836–849, 2015.

[72] C. S. Q. Siew, K. M. Pelczarski, J. S. Yaruss, and M. S. Vitevitch, "Using the OASES-A to illustrate how network analysis can be applied to understand the experience of stuttering," *Journal of Communication Disorders*, vol. 65, Supplement C, pp. 1–9, 2017.

[73] S. Lev-Ari, "The influence of social network size on speech perception," *The Quarterly Journal of Experimental Psychology*, vol. 71, no. 10, pp. 2249–2260, 2018.

[74] S. Lev-Ari, "Social network size can influence linguistic malleability and the propagation of linguistic change," *Cognition*, vol. 176, pp. 31–39, 2018.

[75] L. Steels, "Modeling the cultural evolution of language," *Physics of Life Reviews*, vol. 8, no. 4, pp. 339–356, 2011.

[76] I. T. Koponen and M. Nousiainen, "Concept networks in learning: Finding key concepts in learners' representations of the interlinked structure of scientific knowledge," *Journal of Complex Networks*, vol. 2, no. 2, pp. 187–202, 2014.

[77] I. T. Koponen and M. Pehkonen, "Coherent knowledge structures of physics represented as concept networks in teacher education," *Science and Education*, vol. 19, no. 3, pp. 259–282, 2010.

[78] C. S. Siew, "Using network science to analyze concept maps of psychology undergraduates," *Applied Cognitive Psychology*, pp. 1–7, 2018.

[79] M. Kearns, S. Suri, and N. Montfort, "An experimental study of the coloring problem on human subject networks," *Science*, vol. 313, no. 5788, pp. 824–827, 2006.

[80] W. Mason and D. J. Watts, "Collaborative learning in networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 109, no. 3, pp. 764–769, 2012.

[81] E. A. Karuza, A. E. Kahn, S. L. Thompson-Schill, and D. S. Bassett, "Process reveals structure: How a network is traversed mediates expectations about its architecture," *Scientific Reports*, vol. 7, no. 1, Article ID 12733, 2017.

[82] E. A. Karuza, S. L. Thompson-Schill, and D. S. Bassett, "Local patterns to global architectures: influences of network topology on human learning," *Trends in Cognitive Sciences*, vol. 20, no. 8, pp. 629–640, 2016.

[83] C. T. Butts, "Revisiting the foundations of network analysis," *American Association for the Advancement of Science: Science*, vol. 325, no. 5939, pp. 414–416, 2009.

[84] L. Euler, "Solutio problematis ad geometriam situs pertinensis," *Comm. Acad. Sci. Imper. Petropol*, vol. 8, pp. 128–140, 1736.

[85] S. Boccaletti, G. Bianconi, and R. Criado, "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014.

[86] U. Brandes, S. P. Borgatti, and L. C. Freeman, "Maintaining the duality of closeness and betweenness centrality," *Social Networks*, vol. 44, pp. 153–159, 2016.

[87] D. Koschützki, K. A. Lehmann, L. Peeters et al., "Centrality indices," in *Network analysis: Methodological Foundations*, U. Brandes and T. Erlebach, Eds., pp. 16–61, Springer, 2005.

[88] D. Papo, J. M. Buldú, S. Boccaletti, and E. T. Bullmore, "Complex network theory and the brain," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1653, Article ID 20130520, 2014.

[89] P. A. Luce and D. B. Pisoni, "Recognizing spoken words: the neighborhood activation model," *Ear and Hearing*, vol. 19, no. 1, pp. 1–36, 1998.

[90] M. Coltheart, E. Davelaar, T. Jonasson, and D. Besner, "Access to the internal lexicon," in *Attention and Performance VI*, S. Dornic, Ed., Lawrence Erlbaum Associates, 1977.

[91] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[92] M. Yates, "How the clustering of phonological neighbors affects visual word recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 39, no. 5, pp. 1649–1656, 2013.

[93] K. Y. Chan and M. S. Vitevitch, "Network structure influences speech production," *Cognitive Science*, vol. 34, no. 4, pp. 685–697, 2010.

[94] R. Goldstein and M. S. Vitevitch, "The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition," *Frontiers in Psychology*, vol. 5, no. 1307, 2014.

[95] M. S. Vitevitch, K. Y. Chan, and S. Roodenrys, "Complex network structure influences processing in long-term and short-term memory," *Journal of Memory and Language*, vol. 67, no. 1, pp. 30–44, 2012.

[96] M. A. Beauchamp, "An improved index of centrality," *Behavioural Science*, vol. 10, pp. 161–163, 1965.

[97] R. Goldstein and M. S. Vitevitch, "The influence of closeness centrality on lexical processing," *Frontiers in Psychology*, vol. 8, no. 1683, 2017.

[98] N. Castro and M. Stella, *The Multiplex Structure of The Mental Lexicon Influences Picture Naming in People with Aphasia*, 2018, https://doi.org/10.31234/osf.io/eqvmg.

[99] S. R. S. Iyengar, C. E. V. Madhavan, K. A. Zweig, and A. Natarajan, "Understanding human navigation using network analysis," *Topics in Cognitive Science*, vol. 4, no. 1, pp. 121–134, 2012.

[100] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks*, vol. 30, no. 1, pp. 107–117, 1998.

[101] M. S. Vitevitch, R. Goldstein, and E. Johnson, "Path-length and the misperception of speech: insights from network science and psycholinguistics," in *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, A. Mehler, A. Lücking, and S. Banisch, Eds., pp. 29–45, Springer, Berlin, Germany, 2016.

[102] Y. N. Kenett, E. Levi, D. Anaki, and M. Faust, "The semantic distance task: Quantifying semantic distance with semantic network path length," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 43, no. 9, pp. 1470–1489, 2017.

[103] M. S. Vitevitch, K. Y. Chan, and R. Goldstein, "Insights into failed lexical retrieval from network science," *Cognitive Psychology*, vol. 68, no. 1, pp. 1–32, 2014.

[104] S. P. Borgatti, "Identifying sets of key players in a social network," *Computational and Mathematical Organization Theory*, vol. 12, no. 1, pp. 21–34, 2006.

[105] M. S. Vitevitch and R. Goldstein, "Keywords in the mental lexicon," *Journal of Memory and Language*, vol. 73, no. 1, pp. 131–147, 2014.

[106] C. S. Q. Siew and M. S. Vitevitch, "Spoken word recognition and serial recall of words from components in the phonological network," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 42, no. 3, pp. 394–410, 2016.

[107] M. S. Vitevitch and N. Castro, "Using network science in the language sciences and clinic," *International Journal of Speech-Language Pathology*, vol. 17, no. 1, pp. 13–25, 2015.

[108] C. S. Q. Siew, "The influence of 2-hop network density on spoken word recognition," *Psychonomic Bulletin & Review*, vol. 24, no. 2, pp. 496–502, 2017.

[109] M. S. Vitevitch, R. Goldstein, C. S. Siew, and N. Castro, "Using complex networks to understand the mental lexicon," in *In Yearbook of the Poznan Linguistic Meeting*, vol. 1, pp. 119–138, De Gruyter Open, 2014.

[110] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.

[111] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577–8582, 2006.

[112] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.

[113] E. Bullmore and O. Sporns, "The economy of brain network organization," *Nature Reviews Neuroscience*, vol. 13, no. 5, pp. 336–349, 2012.

[114] C. C. Hilgetag and M.-T. Hütt, "Hierarchical modular brain connectivity is a stretch for criticality," *Trends in Cognitive Sciences*, vol. 18, no. 3, pp. 114-115, 2014.

[115] D. Meunier, R. Lambiotte, and E. T. Bullmore, "Modular and hierarchically modular organization of brain networks," *Frontiers in Neuroscience*, vol. 4, no. 200, 2010.

[116] C. J. Stam, "Modern network science of neurological disorders," *Nature Reviews Neuroscience*, vol. 15, no. 10, pp. 683–695, 2014.

[117] E. C. W. van Straaten and C. J. Stam, "Structure out of chaos: Functional brain network analysis with EEG, MEG, and functional MRI," *European Neuropsychopharmacology*, vol. 23, no. 1, pp. 7–18, 2013.

[118] Y. N. Kenett, R. Gold, and M. Faust, "The Hyper-Modular Associative Mind: A Computational Analysis of Associative Responses of Persons with Asperger Syndrome," *Language and Speech*, vol. 59, no. 3, pp. 297–317, 2016.

[119] Y. N. Kenett, O. Levy, D. Y. Kenett, H. E. Stanley, M. Faust, and S. Havlin, "Flexibility of thought in high creative individuals represented by percolation analysis," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 115, no. 5, pp. 867–872, 2018.

[120] S. Shai, D. Y. Kenett, Y. N. Kenett et al., "Critical tipping point distinguishing two types of transitions in modular network structures," *Physical Review E*, vol. 92, no. 6, Article ID 062805, 2015.

[121] M. Benedek, Y. N. Kenett, K. Umdasch, D. Anaki, M. Faust, and A. C. Neubauer, "How semantic memory structure and intelligence contribute to creative thought: a network science approach," *Thinking and Reasoning*, vol. 23, no. 2, pp. 158–183, 2017.

[122] R. V. Sole and M. Montoya, "Complexity and fragility in ecological networks," in *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 268, pp. 2039–2045, 2001.

[123] M. P. van den Heuvel and O. Sporns, "Network hubs in the human brain," *Trends in Cognitive Sciences*, vol. 17, no. 12, pp. 683–696, 2013.

[124] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, "Tastes, ties, and time: A new social network dataset using Facebook.com," *Social Networks*, vol. 30, no. 4, pp. 330–342, 2008.

[125] R. Albert, H. Jeong, and A.-L. Barabási, "Internet: diameter of the World-Wide Web," *Nature*, vol. 401, no. 6749, pp. 130-131, 1999.

[126] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, pp. 198701–198701-4, 2001.

[127] G. K. Zipf, *Human Behavior and The Principle of Least Effort: An Introduction to Human Ecology*, Addison-Wesley Press, Cambridge, MA, 1949.

[128] J. Goñi, G. Arrondo, J. Sepulcre et al., "The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory," *Cognitive Processing*, vol. 12, no. 2, pp. 183–196, 2011.

[129] A. Clauset, C. R. Shalizi, and M. E. . Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.

[130] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physiscs*, vol. 46, no. 5, pp. 323–351, 2005.

[131] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Resilience of the Internet to random breakdowns," *Physical Review Letters*, vol. 85, no. 21, pp. 4626–4628, 2000.

[132] C. T. Kello, G. D. A. Brown, R. Ferrer-i-Cancho et al., "Scaling laws in cognitive sciences," *Trends in Cognitive Sciences*, vol. 14, no. 5, pp. 223–232, 2010.

[133] A. S. Morais, H. Olsson, and L. J. Schooler, "Mapping the Structure of Semantic Memory," *Cognitive Science*, vol. 37, no. 1, pp. 125–145, 2013.

[134] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[135] A. P. Christensen, "NetworkToolbox: Methods and measures for brain, cognitive, and psychometric network analysis in R," *The R Journal*, vol. 10, pp. 422–439, 2018.

[136] A. P. Christensen, "SemNetCleaner: An automated cleaning tool for semantic and linguistic data," package version 1.0.0, 2019, https://github.com/AlexChristensen/SemNetCleaner.

[137] C. S. Q. Siew, "Spreadr, A R package to simulate spreading activation in a network," *Behavior Research Methods*, pp. 1–20, 2019.

[138] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[139] S. Epskamp, D. Borsboom, and E. I. Fried, "Estimating psychological networks and their accuracy: A tutorial paper," *Behavior Research Methods*, vol. 50, no. 1, pp. 195–212, 2018.

[140] M. R. Quillian, "Word concepts: a theory and simulation of some basic semantic capabilities.," *Behavioural Science*, vol. 12, no. 5, pp. 410–430, 1967.

[141] M. R. Quillian, "The teachable language comprehender: A simulation program and theory of language," *Communications of the ACM*, vol. 12, no. 8, pp. 459–476, 1969.

[142] A. M. Collins and M. R. Quillian, "Retrieval time from semantic memory," *Journal of Verbal Learning and Verbal Behavior*, vol. 8, no. 2, pp. 240–247, 1969.

[143] J. R. Anderson, "A spreading activation theory of memory," *Journal of Verbal Learning and Verbal Behavior*, vol. 22, no. 3, pp. 261–295, 1983.

[144] G. S. Dell, "A Spreading-Activation Theory of Retrieval in Sentence Production," *Psychological Review*, vol. 93, no. 3, pp. 283–321, 1986.

[145] D. A. Balota and R. F. Lorch Jr., "Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 12, no. 3, pp. 336–345, 1986.

[146] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *The New England Journal of Medicine*, vol. 357, no. 4, pp. 370–379, 2007.

[147] T. W. Valente, L. A. Palinkas, S. Czaja, K. H. Chu, and C. H. Brown, "Social network analysis for program implementation," *PloS ONE*, vol. 10, no. 6, Article ID e0131712, 2015.

[148] J. T. Abbott, J. L. Austerweil, and T. L. Griffiths, "Random walks on semantic networks can resemble optimal foraging," *Psychological Review*, vol. 122, no. 3, pp. 558–569, 2015.

[149] D. D. Bourgin, J. T. Abbott, T. L. Griffiths, K. A. Smith, and E. Vul, "Empirical evidence for markov chain monte carlo in memory search," in *Proceedings of the In Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pp. 224–229, Boston, MA, USA, 2014.

[150] M. I. Fathan, E. K. Renfro, J. L. Austerweil, and N. M. Beckage, "Do Humans Navigate via Random Walks? Modeling Navigation in a Semantic Word Game. Cognitive Science Conference," in *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, T. T. Rogers, M. Rau, X. Zhu, and C. W. Kalish, Eds., pp. 366–371, Austin, TX, USA, 2018.

[151] K. A. Smith, D. E. Huber, and E. Vul, "Multiply-constrained semantic search in the Remote Associates Test," *Cognition*, vol. 128, no. 1, pp. 64–75, 2013.

[152] M. S. Vitevitch, G. Ercal, and B. Adagarla, "Simulating retrieval from a highly clustered network: Implications for spoken word recognition," *Frontiers in Psychology*, vol. 2, Article ID 369, 2011.

[153] Y. N. Kenett and M. Faust, "A semantic network cartography of the creative mind," *Trends in Cognitive Sciences*, vol. 23, no. 4, pp. 274–276, 2019.

[154] S. Mednick, "The associative basis of the creative process," *Psychological Review*, vol. 69, no. 3, pp. 220–232, 1962.

[155] E. Volle, "Associative and controlled cognition in divergent thinking: Theoretical, experimental, neuroimaging evidence, and new directions," in *The Cambridge Handbook of the Neuroscience of Creativity*, R. E. Jung and O. Vartanian, Eds., pp. 333–360, Cambridge University Press, New York, NY, USA, 2018.

[156] Y. N. Kenett and J. L. Austerweil, "Examining search processes in low and high creative individuals with random walks," in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, A. Papafragou D and J. C. Trueswell, Eds., pp. 313–318, Cognitive Science Society, Austin, TX, USA, 2016.

[157] K. Gray, S. Anderson, E. Chen et al., "Forward flow: A new measure to quantify free thought and predict creativity," *American Psychologist*, 2019.

[158] Y. N. Kenett, "Investigating creativity from a semantic network perspective," in *Exploring Transdisciplinarity in Art and Science*, Z. Kapoula, E. Volle, J. Renoult, and M. Andreatta, Eds., pp. 49–75, Springer, 2018.

[159] T. T. Hills, P. M. Todd, and R. L. Goldstone, "Search in external and internal spaces: Evidence for generalized cognitive search processes," *Psychological Science*, vol. 19, no. 8, pp. 802–808, 2008.

[160] T. T. Hills, M. N. Jones, and P. M. Todd, "Optimal foraging in semantic memory," *Psychological Review*, vol. 119, no. 2, pp. 431–440, 2012.

[161] D. U. Wulff, T. T. Hills, and R. Hertwig, "Worm holes in memory: Is memory one representation or many?" in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, pp. 3817–3822, Cognitive Science Society, 2013.

[162] T. T. Hills, "Animal foraging and the evolution of goal-directed cognition," *Cognitive Science*, vol. 30, no. 1, pp. 3–41, 2006.

[163] N. Beckage, M. Steyvers, and C. Butts, "Route choice in individualssemantic network navigation," in *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, N. Miyake, D. Peebles, and R. Cooper, Eds., pp. 108–113, Cognitive Science Society, Austin, TX, USA, 2012.

[164] J. M. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, no. 6798, p. 845, 2000.

[165] R. West and J. Leskovec, "Human wayfinding in information networks," in *Proceedings of the 21st Annual Conference on World Wide Web, WWW'12*, pp. 619–628, ACM, April 2012.

[166] R. West, J. Pineau, and D. Precup, "Wikispeedia: An online game for inferring semantic distances between concepts," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI-09*, pp. 1598–1603, USA, July 2009.

[167] W. A. Bousfield and C. H. W. Sedgewick, "An analysis of sequences of restricted associative responses," *The Journal of General Psychology*, vol. 30, no. 2, pp. 149–165, 1944.

[168] A. K. Romney, D. D. Brewer, and W. H. Batchelder, "Predicting clustering from semantic structure," *Psychological Science*, vol. 4, no. 1, pp. 28–34, 1993.

[169] Y. N. Kenett, D. Wechsler-Kashi, D. Y. Kenett, R. G. Schwartz, E. Ben-Jacob, and M. Faust, "Semantic organization in children with cochlear implants: Computational analysis of verbal fluency," *Frontiers in Psychology*, vol. 4, Article ID 543, pp. 1–11, 2013.

[170] D. U. Wulff, T. T. Hills, M. Lachman, and R. Mata, "The aging lexicon: Differences in the semantic networks of younger and older adults," in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, A. Papafragou D and J. C. Trueswell, Eds., pp. 313–318, Austin, TX, USA, 2016.

[171] Y. N. Kenett, R. E. Beaty, P. J. Silvia, D. Anaki, and M. Faust, "Structure and flexibility: Investigating the relation between the structure of the mental lexicon, fluid intelligence, and creative achievement," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 10, no. 4, pp. 377–388, 2016.

[172] K. Borodkin, Y. N. Kenett, M. Faust, and N. Mashal, "When pumpkin is closer to onion than to squash: The structure of the second language lexicon," *Cognition*, vol. 156, pp. 60–70, 2016.

[173] A. P. Christensen, Y. N. Kenett, K. N. Cotter, R. E. Beaty, and P. J. Silvia, "Remotely Close Associations: Openness to Experience and Semantic Memory Structure," *European Journal of Personality*, vol. 32, no. 4, pp. 480–492, 2018.

[174] A. K. Troyer, M. Moscovitch, and G. Winocur, "Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults," *Neuropsychology*, vol. 11, no. 1, pp. 138–146, 1997.

[175] T. Hills, R. Mata, A. Wilke, and G. Samanez-Larkin, "Exploration and exploitation in memory search across the lifespan,"

in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 33, 2011.

[176] T. T. Hills and T. Pachur, "Dynamic search and working memory in social recall," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 38, no. 1, pp. 218–228, 2012.

[177] J. E. Avery and M. N. Jones, "Comparing models of semantic fluency: Do humans forage optimally or walk randomly?" in *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, T. T. Rogers, M. Rau, X. Zhu, and C. W. Kalish, Eds., pp. 118–123, Austin, TX, USA, 2018.

[178] N. Unsworth, "Examining the dynamics of strategic search from long-term memory," *Journal of Memory and Language*, vol. 93, pp. 135–153, 2017.

[179] T. T. Hills, P. M. Todd, and M. N. Jones, "Foraging in Semantic Fields: How We Search Through Memory," *Topics in Cognitive Science*, vol. 7, no. 3, pp. 513–534, 2015.

[180] M. N. Jones, T. T. Hills, and P. M. Todd, "Hidden processes in structural representations: A reply to Abbott, Austerweil, and Griffiths (2015)," *Psychological Review*, vol. 122, no. 3, pp. 570–574, 2015.

[181] A. Nematzadeh, F. Miscevic, and S. Stevenson, "Simple search algorithms on semantic networks learned from language use, 2016," https://arxiv.org/abs/1602.03265.

[182] M. N. Jones and D. J. K. Mewhort, "Representing word meaning and order information in a composite holographic lexicon," *Psychological Review*, vol. 114, no. 1, pp. 1–37, 2007.

[183] Y. N. Kenett and S. L. Thompson-Schill, "Dynamic effects of conceptual combination on semantic network structure," in *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, G. Gunzeimann, A. Howes, T. Tenbrinck, and E. Davelaar, Eds., pp. 657–662, Austin, TX, USA, 2017.

[184] L. Barceló-Coblijn, B. Corominas-Murtra, and A. Gomila, "Syntactic trees and small-world networks: Syntactic development as a dynamical process," *Adaptive Behavior*, vol. 20, no. 6, pp. 427–442, 2012.

[185] H. Chen, X. Chen, and H. Liu, "How does language change as a lexical network? An investigation based on written Chinese word co-occurrence networks," *PLoS ONE*, vol. 13, no. 2, Article ID e0192545, 2018.

[186] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition?" *Psychological Science*, vol. 20, no. 6, pp. 729–739, 2009.

[187] L. Fenson, E. Bates, and P. S. Dale, *Macarthur-Bates Communicative Development Inventories*, H. Paul, Ed., Brookes Publishing Company, 2007.

[188] N. Beckage, L. Smith, and T. Hills, "Small worlds and semantic network growth in typical and late talkers," *PLoS ONE*, vol. 6, no. 5, Article ID e19348, 2011.

[189] N. M. Beckage, A. Aguilar, and E. Colunga, "Modeling lexical acquisition through networks," in *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, TX, USA, 2015.

[190] N. M. Beckage and E. Colunga, "Using the words toddlers know now to predict the words they will learn next," in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, TX, USA, 2013.

[191] T. T. Hills, J. Maouene, B. Riordan, and L. B. Smith, "The associative structure of language: Contextual diversity in early word learning," *Journal of Memory and Language*, vol. 63, no. 3, pp. 259–273, 2010.

[192] T. T. Hills and C. S. Q. Siew, "Filling gaps in early word learning," *Nature Human Behavior*, vol. 2, pp. 662-663, 2018.

[193] M. T. Carlson, M. Sonderegger, and M. Bane, "How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions," *Journal of Memory and Language*, vol. 75, pp. 159–180, 2014.

[194] H. L. Storkel, "Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants," *Journal of Child Language*, vol. 36, no. 2, pp. 291–321, 2009.

[195] T. Engelthaler and T. T. Hills, "Feature Biases in Early Word Learning: Network Distinctiveness Predicts Age of Acquisition," *Cognitive Science*, vol. 41, pp. 120–140, 2017.

[196] A. E. Sizemore, E. A. Karuza, C. Giusti, and D. S. Bassett, "Knowledge gaps in the early growth of semantic networks," *Nature Human Behaviour*, vol. 2, no. 9, pp. 682–692, 2018.

[197] M. Zortea, B. Menegola, A. Villavicencio, and J. F. D. Salles, "Graph analysis of semantic word association among children, adults, and the elderly," *Psicologia: Reflexao e Critica*, vol. 27, no. 1, pp. 90–99, 2014.

[198] M. Karl Healey and M. J. Kahana, "A four-component model of age-related memory change," *Psychological Review*, vol. 123, no. 1, pp. 23–69, 2016.

[199] T. A. Salthouse, "Selective review of cognitive aging," *Journal of the International Neuropsychological Society*, vol. 16, no. 5, pp. 754–760, 2010.

[200] M. Ramscar, C. C. Sun, P. Hendrix, and R. H. Baayen, "The Mismeasurement of Mind: Life-Span Changes in Paired-Associate-Learning Scores Reflect the "Cost" of Learning, Not Cognitive Decline," *Psychological Science*, vol. 28, no. 8, pp. 1171–1179, 2017.

[201] J. Borge-Holthoefer, Y. Moreno, and A. Arenas, "Modeling abnormal priming in Alzheimers patients with a free association network," *PloS ONE*, vol. 6, no. 8, Article ID e22651, 2011.

[202] M. Laisney, B. Giffard, S. Belliard, V. de la Sayette, B. Desgranges, and F. Eustache, "When the zebra loses its stripes: Semantic priming in early Alzheimer's disease and semantic dementia," *Cortex*, vol. 47, no. 1, pp. 35–46, 2011.

[203] M. A. Schilling, "A "small-world" network model of cognitive insight," *Creativity Research Journal*, vol. 17, no. 2-3, pp. 131–154, 2005.

[204] J. C. Zemla, Y. N. Kenett, K.-S. Jun, and J. L. Austerweil, "U-INVITE: Estimating individual semantic networks from fluency data," in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, A. Papafragou D and J. C. Trueswell, Eds., pp. 1907–1912, Austin, TX, USA, 2016.

[205] S. Moreno and J. Neville, "Network hypothesis testing using mixed kronecker product graph models," in *Proceedings of the 13th IEEE International Conference on Data Mining, ICDM 2013*, pp. 1163–1168, USA, December 2013.

[206] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Bootstrap percolation on complex networks," *Physical Review E*, vol. 82, no. 1, Article ID 011103, 2010.

[207] C. S. Q. Siew and M. S. Vitevitch, "The phonographic language network: Using network science to investigate the phonological and orthographic similarity structure of language," *Journal of Experimental Psychology: General*, pp. 1–25, 2019.

[208] J. O. Garcia, A. Ashourvan, S. Muldoon, J. M. Vettel, and D. S. Bassett, "Applications of community detection techniques to brain graphs: algorithmic considerations and implications for neural function," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 846–867, 2018.

[209] Y. N. Kenett, J. D. Medaglia, R. E. Beaty et al., "Driving the brain towards creativity and intelligence: A network control theory analysis," *Neuropsychologia*, vol. 118, pp. 79–90, 2018.

[210] J. D. Medaglia, "Clarifying cognitive control and the controllable connectome," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 10, no. 1, Article ID e1471, 2019.

[211] E. Tang and D. S. Bassett, "Colloquium: Control of dynamics in brain networks," *Reviews of Modern Physics*, vol. 90, no. 3, 2018.

Hindawi
Complexity
Volume 2019, Article ID 4203158, 12 pages
https://doi.org/10.1155/2019/4203158

WILEY | Hindawi

Research Article

Analyzing Knowledge Retrieval Impairments Associated with Alzheimer's Disease Using Network Analyses

Jeffrey C. Zemla and Joseph L. Austerweil

Department of Psychology, University of Wisconsin-Madison, USA

Correspondence should be addressed to Jeffrey C. Zemla; zemla@wisc.edu

Received 2 September 2018; Revised 19 February 2019; Accepted 19 March 2019; Published 2 May 2019

Guest Editor: Dirk Wulff

A defining characteristic of Alzheimer's disease is difficulty in retrieving semantic memories, or memories encoding facts and knowledge. While it has been suggested that this impairment is caused by a degradation of the semantic store, the precise ways in which the semantic store is degraded are not well understood. Using a longitudinal corpus of semantic fluency data (listing of items in a category), we derive semantic network representations of patients with Alzheimer's disease and of healthy controls. We contrast our network-based approach with analyzing fluency data with the standard method of counting the total number of items and perseverations in fluency data. We find that the networks of Alzheimer's patients are more connected and that those connections are more randomly distributed than the connections in networks of healthy individuals. These results suggest that the semantic memory impairment of Alzheimer's patients can be modeled through the inclusion of spurious associations between unrelated concepts in the semantic store. We also find that information from our network analysis of fluency data improves prediction of patient diagnosis compared to traditional measures of the semantic fluency task.

1. Introduction

Alzheimer's disease (AD) is a debilitating neurodegenerative disease that affects roughly 46 million people worldwide [1]. A defining characteristic of AD is an increased difficulty in retrieving semantic memories (i.e., declarative knowledge of facts and concepts). Patients with AD have difficulty naming objects [2], matching semantically related pictures [3], and identifying the semantic features of words [4]. Though there are many neuropsychological tests for measuring semantic impairment due to AD, the mechanisms producing these deficits have not been isolated. While some research suggests these deficits are due to a degradation of a semantic memory store that encodes concepts in the mind [5–7], other evidence points to difficulty in retrieving memories from an intact semantic memory store [8–10].

One difficulty in resolving this debate is that the semantic memory store is not directly observable. To address this, computational models have been developed to explain how semantic memories are represented and to make inferences about the underlying mechanisms responsible for memory retrieval [11]. However most techniques for estimating semantic representations assume a common knowledge representation for a group, including patient populations [5, 12]. This is problematic for analyzing individuals with AD as their impairments tend to be heterogeneous [13, 14]. Aggregating over retrieval data of many patients to estimate a single group-based representation may result in an estimated representation that does not actually resemble any individual in the population [15, 16]. In this article, we use semantic networks as a model of how memories of facts and knowledge are encoded, develop a method for estimating networks from memory retrieval data, and use it to analyze data from individuals with AD and healthy controls at individual and cross-sectional levels.

There is a long history of modeling semantic knowledge using a semantic network [17, 18], an abstract representation of how concepts are organized in the mind. In a semantic network, concepts are represented by nodes and semantic similarity is represented by edges that connect pairs of nodes. In recent years, advances in network science have improved our understanding of semantic memory by providing us

tools to quantify how networks are organized [19], and as a result semantic networks have been used to explore how conceptual knowledge is affected by factors such as creativity [20], bilingualism [21], age [22], and more.

Among the most commonly used tasks to diagnose semantic memory impairment is the semantic fluency task [23], in which participants list as many items from a category (e.g., animals) as they can in a short period of time (e.g., one to three minutes). This test is part of several popular neuropsychological batteries, including the Cognitive Linguistic Quick Test [24] and Uniform Data Set [25]. Traditionally, the fluency task is scored by counting the number of perseverations (repetitions) listed by a participant, as well as the number of items listed, excluding perseverations and errors (responses not in the target category). Compared to healthy controls, individuals with AD routinely list fewer items [26] and have higher perseveration rates [27]. Even in presymptomatic individuals, perseverations in the fluency task have been associated with future cognitive decline [28], and the number of items listed has been associated with pathological markers of AD [29].

However semantic fluency data can also be used to estimate semantic networks of groups or individuals [30, 31]. This is possible because semantic fluency data tend to be clustered [32]: individuals often list multiple semantically related responses in sequence. For instance, when listing animals, a participant may list a sequence of pets (such as *dog, cat*, and *hamster*) before switching to a new cluster (e.g., zoo animals such as *giraffe, lion*, and *hippo*). Because semantically related words typically appear near each other in a fluency list, a semantic network of word associations can be estimated from a corpus of fluency data.

In this article, we apply a random walk model of semantic memory retrieval [31, 33] to a longitudinal dataset of semantic fluency data from AD patients and healthy controls in order to estimate semantic network representations of individuals and investigate mechanisms responsible for impaired performance due to AD. We compare these representations and find systematic differences in the structure of semantic representations between AD patients and control participants. This network-based analysis of semantic fluency data provides additional insight into the specific cognitive mechanisms that lead to memory impairments by identifying associations between properties of an individual's semantic network and impaired behavioral performance.

To model impaired performance, we extend the random walk model of semantic memory retrieval [33] to account for perseverations in semantic fluency data. Perseverations in semantic fluency data are modeled as errors resulting from a faulty monitoring process associated with working memory, which is in line with current neuropsychiatric research [34]. We propose a generative computational model that accounts for perseverations in fluency data and demonstrate that it can quantitatively capture the severity of impairment to this monitoring process.

2. Materials and Methods

2.1. Participants and Design. We acquired a longitudinal corpus of semantic fluency data from the University of California-San Diego Shiley-Marcos Alzheimer's Disease Research Center (ADRC). Data were collected between 1985 and 2016 as part of the ADRC's broader goal to better understand Alzheimer's disease. The fluency data used for our analyses partially overlaps with data presented in previous publications from this ADRC (e.g., [35]).

Each participant visited the lab approximately once per year for the duration of their involvement and was tested on the semantic fluency task as part of a longer neuropsychiatric exam. The average length of participation was approximately 9 years per participant (i.e., most participants began after 1985 and/or discontinued participation prior to 2016.) Participants included healthy individuals, as well as those who were already diagnosed with AD or other memory-related issues. At each visit, patients were given a clinical diagnosis using the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) scale [36]. This evaluation was based on multiple sources, including the participant's performance on the Mini-Mental State Exam [37]. Of interest, these diagnoses included "Normal Control" (NC) and "Probable AD" (PAD), but a much smaller number of participants were given other diagnoses (such as "Mild Cognitive Impairment" or "Frontotemporal dementia"). To focus our analyses on AD and not related dementias, we limit our analysis of the dataset only to visits in which a participant was diagnosed as NC or PAD; all other visits were excluded. Diagnoses of AD using the NINCDS-ADRDA scale have been found to have good sensitivity and specificity when compared to postmortem pathological reports [38]; however, the scale is limited in that it does not account for secondary diagnoses (e.g., some patients diagnosed as PAD may also have pathological signs of Lewy body dementia). Newer clinical definitions of Alzheimer's have evolved over time, most notably making use of in vivo biomarker detection for diagnosis (e.g., [39]). However due to the longitudinal nature of the dataset, a fixed classification scheme was used.

The animal semantic fluency task lasted one minute per visit. Participants named animals aloud, which were written down in real time on paper by a researcher conducting the task. In total, we transcribed 1,047 animal fluency lists generated from 123 participants (60% female, mean age at first visit 71.4, range 34–90). This set of participants excludes any individual who did not have at least three fluency lists when diagnosed as NC or three fluency lists when diagnosed as PAD. These lists are nonoverlapping. For example, a participant with 2 NC and 2 PAD visits was rejected outright (not included in the 123 participants). A participant with 3 NC and 1 PAD visit would have their NC visits analyzed, but not their PAD visits. Of those, 248 lists were generated from 41 participants who were diagnosed as PAD for that visit (51% female, mean age 75.2, range 61–90), while 799 lists were generated from 84 participants who were diagnosed as NC for that visit (64% female, mean age 69.5, range 34–86). These participant pools overlapped—2 participants were diagnosed as both NC and PAD (on separate visits).

List 1: hamster, dog, cat, lion, …
List 2: lion, elephant, ~~lion~~, cat, hamster, …
List 3: cat, leopard, zebra, ~~leopard~~, ~~cat~~, dog,…

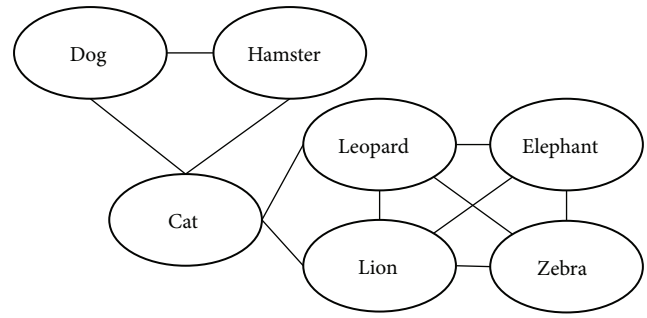


Figure 1: Semantic fluency lists (left) can be modeled as a censored random walk on a semantic network (right). When $p_{emit} = 0$, repeated items are "censored" on subsequent traversals, as shown above. When $0 < p_{emit} < 1$, this censoring process is stochastic. Figure reprinted from Zemla and Austerweil [31] with permission from Springer.

2.2. Network Estimation Model. As suggested by current best practices for estimating undirected, unweighted semantic networks from fluency data [31], we used U-INVITE to infer networks for each individual from their fluency data. U-INVITE is a method for estimating networks which assumes fluency data are generated by a censored random walk on that network [33, 40]. Assuming an individual's fluency data are generated by a censored random walk, U-INVITE uses Bayesian inference to estimate the most likely network. In a censored random walk, states in the walk are observed when they are traversed for the first time, but are "censored" (unobserved) on subsequent traversals. For example, if a random walk on a network produces the list "dog, cat, hamster, cat, lion," the censored list would be "dog, cat, hamster, lion"—the second occurrence of "cat" is censored (see Figure 1). This model has been shown to approximate human fluency data in many ways [33, 41]. As previous work focused on healthy individuals, the censoring process was deterministic and did not produce perseverations (repeated items). Repeating items during the semantic fluency task is a hallmark of Alzheimer's fluency data. To account for perseverations, we modify this process so that data are generated by a noisy censored random walk: repeated items are emitted with some unknown probability p_{emit} and are censored with probability $1-p_{emit}$. When $p_{emit} = 0$, censoring is deterministic (i.e., no items are ever repeated in the censored random walk) and it is equivalent to the previous model.

Under this model, the probability of a semantic network given a set of L fluency lists $X = \{X^1, \ldots, X^L\}$ is

$$\begin{aligned}&\mathbb{P}\left(\mathbf{G} \mid X^1, \ldots, X^L, p_{emit}\right)\\ &\quad\propto \mathbb{P}\left(\mathbf{G}\right)\mathbb{P}\left(p_{emit}\right)\prod_{l=1}^{L}\mathbb{P}\left(X^l \mid \mathbf{G}, p_{emit}\right)\end{aligned} \tag{1}$$

where $\mathbf{G}$ denotes an undirected and unweighted network, and G_{ij} (for each i and j) is either 1 or 0 to indicate whether an edge exists between the two concepts associated with indices i and j. The likelihood of generating any fluency list given a network is the product of all transition probabilities in that list, multiplied by the probability of observing the initial item in that list

$$\begin{aligned}&\mathbb{P}\left(X^l \mid \mathbf{G}, p_{emit}\right)\\ &\quad= \mathbb{P}\left(X_1^l \mid \mathbf{G}\right)\prod_{n=2}^{N_l}\mathbb{P}\left(X_n^l \mid X_1^l, \ldots, X_{n-1}^l, \mathbf{G}, p_{emit}\right)\end{aligned} \tag{2}$$

where X_n^l denotes the nth item of the lth fluency list, and N_l denotes the number of items in the lth list. We assume that the probability of the initial item in a list is given by the limiting probability of an infinite-length random walk encountering that item's node (the stationary distribution of a random walk over the network):

$$\mathbb{P}\left(X_1^l = i \mid \mathbf{G}\right) = \frac{\sum_{m=1}^{M} G_{im}}{\sum_{m=1}^{M}\sum_{p=1}^{M} G_{pm}} \tag{3}$$

where M denotes the number of nodes in network $\mathbf{G}$ (i.e., the total number of unique responses across all lists in X). In other words, the probability of an initial item in a list is proportional to the number of edges connected to that item in $\mathbf{G}$. In (3) and elsewhere, we use the subscript of an item label and its index within a matrix interchangeably (i.e., if $X_1^l = i$ = "dog", then G_{im} indicates whether an edge exists between "dog" and the item label associated with index m in $\mathbf{G}$).

Each transition probability can be modeled as an absorbing random walk. First, we translate link matrix $\mathbf{G}$ into a transition probability matrix $\mathbf{A}$, where

$$A_{ij} = \frac{G_{ij}}{\sum_{m=1}^{M} G_{im}} \tag{4}$$

We rearrange the rows and columns of $\mathbf{A}$ to be in list order (the same order as X^l), which we denote as $\mathbf{A}^l$. Items that do not appear in X^l are excluded from $\mathbf{A}^l$. When perseverations occur in X^l, only the first occurrence is preserved in $\mathbf{A}^l$ (so that each node appears at most once in $\mathbf{A}^l$).

For each transition probability $\mathbb{P}(X_n^l \mid X_1^l, \ldots, X_{n-1}^l, \mathbf{G}, p_{emit})$ that is calculated, $\mathbf{A}^l$ is decomposed into submatrices:

$$\mathbf{A}^l = \begin{bmatrix}\mathbf{Q} & \mathbf{R}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \tag{5}$$

where $\mathbf{Q}$ denotes transitions between nodes observed prior to the currently considered transition to node n (i.e., nodes

in $\{X_1^l, \ldots, X_{n-1}^l\}$) and $\mathbf{R}$ denotes transitions from previously observed nodes to new nodes (i.e., nodes in $\{X_n^l, \ldots, X_{N_l}^l\}$). 0 and $\mathbf{I}$ denote a matrix of zeroes and the identity matrix, respectively. We then define $\mathbf{Q}'$ as

$$\mathbf{Q}' = (1 - p_{emit}) \cdot \mathbf{Q} \tag{6}$$

and $\mathbf{R}'$ as

$$\mathbf{R}' = [p_{emit} \cdot \mathbf{Q}, \mathbf{R}] \tag{7}$$

$\mathbf{Q}'$ denotes the probabilities of transitioning from any previously observed node to another previously observed node *while being censored*. $\mathbf{R}'$ denotes the probabilities of transitioning from any previously observed node to *either* a new node *or* a previously observed node that is not censored. While $\mathbf{Q}'$ is of the same dimension as $\mathbf{Q}$, $\mathbf{R}'$ is larger than $\mathbf{R}$: $\mathbf{R}'$ contains the same number of rows as $\mathbf{R}$, but the number of columns in $\mathbf{R}'$ is equal to the total number of unique items in X^l.

We can then calculate a transition probability as

$$\mathbb{P}\left(X_n^l = i \mid X_1^l, \ldots, X_{n-1}^l = j, \mathbf{G}, p_{emit}\right) = \begin{cases} \sum_{k=1}^{s} E_{jk} R'_{ki} & \text{if } \mathbf{E} \text{ exists} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where s denotes the number of unique items listed prior to X_n^l (i.e., the number of rows in $\mathbf{Q}'$). $\mathbf{E}$ is the fundamental matrix of the Markov chain for transition n [42]:

$$\mathbf{E} = \left(\mathbf{I} - \mathbf{Q}'\right)^{-1} \tag{9}$$

and E_{jk} denotes the expected number of times a Markov chain starting at node j in transition matrix $\mathbf{Q}'$ will visit k before being absorbed.

We derive the prior probability of a network $\mathbb{P}(\mathbf{G})$ using an unweighted and undirected semantic network constructed from the free association norms compiled by the University of South Florida (USF; [43]). These norms were generated by asking over 6,000 participants to respond to a set of cue words with the first meaningfully related word; for instance, if the cue word is "car", a participant might respond "road". From these norms, we extracted all animal cue-response pairs (e.g., "dog–cat") and constructed a semantic network by adjoining each of these pairs with an edge. The network consists of 160 animals and 393 edges.

We assume that the prior probability of an edge in a network is binomial distributed according to whether it occurs in the USF network: $\mathbb{P}(G_{ij} = 1) = 2/3$ when an edge exists between i and j in the USF network, $\mathbb{P}(G_{ij} = 1) = .4$ when an edge does not exist in the USF network, and $\mathbb{P}(G_{ij} = 1) = .5$ when either i or j (or both) are not present in the USF network. These free parameters were derived by using a zero-inflated beta-binomial prior, as described in the hierarchical model of Zemla and Austerweil [31], but treating the USF network as the sole, fixed prior network. As such, $\mathbb{P}(\mathbf{G}) = \prod_{i,j} \mathbb{P}(G_{ij})$ for all i and j in $\mathbf{G}$. Given the data, we seek to find the network that maximizes the *a posteriori* probability:

$$\arg\max_{\mathbf{G}, p_{emit}} \mathbb{P}\left(\mathbf{G}, p_{emit} \mid X\right) \tag{10}$$

We do this through stochastic search on the network. We randomly toggle an edge in the network and accept that edge change when the posterior probability of the network after the edge change is greater than the posterior probability of the network before the edge change. We use a set of heuristics to decide which edges to flip and set a tolerance value such that the network "converges" after 300 edge flips that do not increase $\mathbb{P}(\mathbf{G}, p_{emit} \mid X)$. For further details, see Zemla and Austerweil [31]. After each successful edge toggle, we perform a grid search to find the optimal value for $p_{emit} \in \{0.0, 0.01, \ldots, 0.99, 1.0\}$ given that network. We assume the prior probability of p_{emit} is uniformly distributed over these values.

2.3. Participant Networks and Mock Networks. We estimated a semantic network for each participant by diagnosis (NC or PAD) combination in the data set. Each participant had at least three fluency lists available to generate a network. In total, 125 semantic networks were generated: 84 NC networks and 41 PAD networks. This includes two participants who transitioned from healthy to Alzheimer's diagnosis and had both an NC and PAD network. (More than 2 participants in the dataset converted from NC to PAD, but only 2 participants had a minimum of three NC and three PAD lists required to estimate both networks). The remaining participants (82 NC and 39 PAD) had only one network. PAD networks were generated from an average of 6.05 lists per network (range 3–10), while NC networks were generated from an average of 9.51 lists per network (range 3–26).

We compared PAD networks to NC networks using the following network measures: number of nodes, diameter, density, mean/median node degree, average shortest-path length, clustering coefficient, and small-world coefficient. These measures are further defined in Table 1.

Participants with more fluency lists (and longer fluency lists) will typically have semantic networks that have more nodes and more edges. This is confounding because most network properties (such as diameter or average shortest-path length) are affected by the number of nodes and edges in a network. This makes it difficult to draw inferences from a direct comparison of NC and PAD networks, as the networks vary in the amount of data used to generate them.

To alleviate this problem, we analyzed each network by comparing it to its own set of mock networks generated by the following procedure: For each participant, we generated a random permutation of each fluency list so that the order of the words in each list is arbitrary. We then estimated a network for the set of permuted lists using the same process as their actual network. We repeated this procedure fifty times for each participant. This process ensures that each mock network has the same number of nodes as its corresponding participant network. It also ensures that differences between individuals are not merely due to differences in

Table 1: Network measures.

Measure	Definition
Number of nodes	The total number of nodes in a network
Diameter	The longest shortest-path between any two nodes in a network
Density	A ratio of the number of edges in a network compared to the total number of possible edges in that network
Average shortest-path length	The average length of the shortest-paths between all pairs of nodes
Clustering coefficient	A measure of a network's tendency for a node's neighbors to be connected to each other, defined as 3 times the number of triangles over the number of connected triplets [44]
Small-world coefficient	A measure of a network's "small-worldness" [45]. A small-world network refers to one that has a high clustering coefficient but low average shortest-path length
Node degree	The number of edges connected to a node. Mean degree is the average of every node's degree in a network

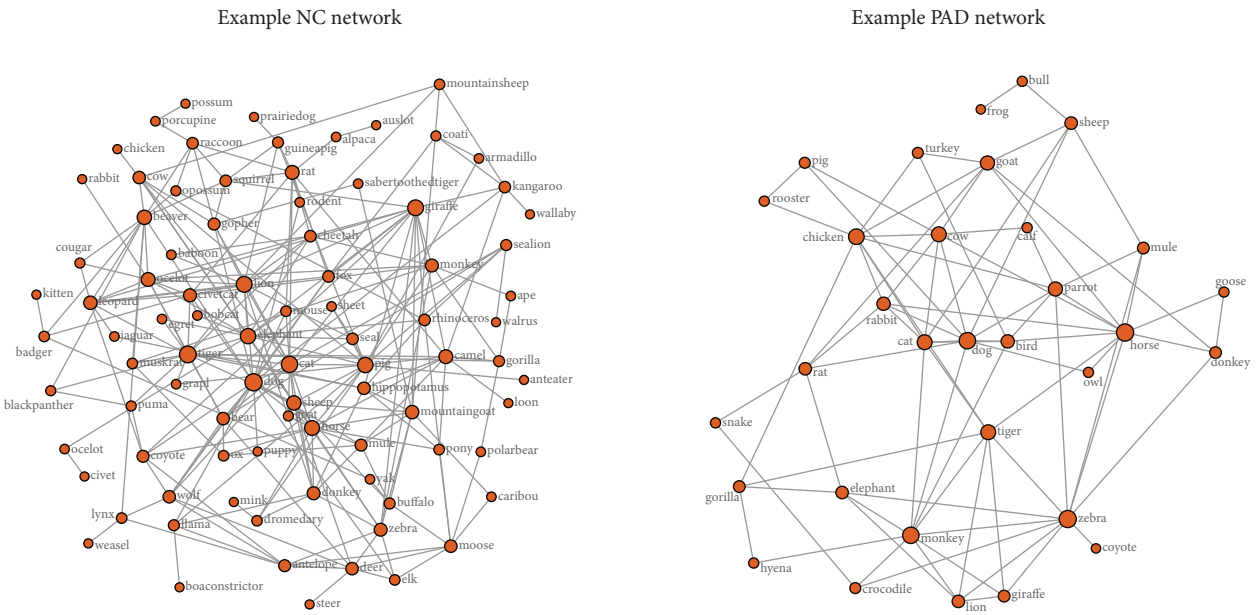


Figure 2: An example network is shown for one NC participant and one PAD participant.

the distribution of animal frequencies in their lists. Using these mock networks, we can define a distribution of values for any network measure under the assumption that words within a list are arbitrarily ordered. Using this bootstrapping procedure, we can then use standard hypothesis testing techniques to gauge how a participant's semantic network deviates from other possible networks that could have been inferred (with the exact same amount of fluency data).

3. Results and Discussion

3.1. Semantic Network Properties and Model Parameters. Estimated semantic networks are available as Supplementary Material (available here). An example semantic network for one PAD participant and one NC participant is shown in Figure 2.

PAD patients listed fewer items per list compared to NC patients, $M_{NC} = 19.33$, $M_{PAD} = 13.13$, $t(123) = 9.76$, $p < .001$, and also had higher rates of perseveration per list, $M_{NC} = .034$, $M_{PAD} = .127$, $t(42.9) = 6.92$, $p < .001$. We applied a Welch correction here and throughout the paper whenever variances were deemed unequal by an F-test. None of these corrections changes the significance of the test (i.e., they did not affect a decision to reject the null hypothesis).

For descriptive purposes, we present a raw comparison between PAD and NC networks without adjustment using the mock networks. (Many of the factors we examine have some degree of correlation. A full correlation matrix between factors is provided in the Supplementary Material). A summary of the network properties for each network type is shown in Table 2. PAD networks appear different than the NC networks in many ways. On average, NC networks have more nodes than PAD networks, a reflection of the fact that PAD patients list fewer unique animals than NC participants, $M_{NC} = 66.8, M_{PAD} = 32.3, t(123.0) = 10.81, p < .001$. In contrast, PAD networks are denser, $M_{NC} = .060, M_{PAD} = .116, t(47.0) = 7.08, p < .001$, and have a smaller diameter, $M_{NC} = 9.55, M_{PAD} = 7.61, t(123) = 3.34, p < .001$.

Table 2: Summary statistics for both PAD an NC networks, as well as mock PAD and NC networks. A dashed line indicates no difference between the mock networks and nonmock statistic. Average shortest-path length and diameter were computed on the largest component of each network, as they are undefined on networks with multiple components.

	NC	NC_{mock}	PAD	PAD_{mock}
Number of networks	84	—	41	—
Number of lists	9.51	—	6.05	—
Number of items listed	19.3	—	13.1	—
Number of nodes	66.8	—	32.3	—
Mini-mental state exam (MMSE)	29.2	—	22.4	—
Diameter	9.55	7.12	7.61	6.37
Density	.06	.07	.12	.13
Mean node degree	3.55	4.37	3.18	3.44
Median node degree	2.55	2.93	2.37	2.51
Average shortest-path length	3.74	3.07	3.25	2.89
Clustering coefficient	.12	.15	.14	.17
Perseveration rate	.034	—	.127	—
Perseveration parameter (p_{emit})	.071	.097	.345	.304
Small-world coefficient	1.92	2.05	1.21	1.36

The increased density of PAD networks could reflect an increase in the number of spurious associations. This is consistent with previous behavioral findings; for instance, Chan, Butters, Salmon, and McGuire [46] found that while a cohort of AD patients were unimpaired matching animal names to pictures, they tended to group animals into atypical categories.

Perhaps because they are more dense, PAD networks tend to have a shorter average shortest-path length, $M_{NC} = 3.74, M_{PAD} = 3.25, t(123) = 3.24, p = .002$. NC networks have a higher mean degree, $M_{NC} = 3.55, M_{PAD} = 3.18, t(123) = 2.69, p = .008$, meaning that healthy control networks have, on average, more semantic associates per concept than AD networks. AD and NC networks do not differ in their median degree ($p = .18$) or in their clustering coefficients ($p = .15$).

A small-world network is one that has a small average shortest-path length but high clustering coefficient [47]. Small-world networks are efficient in that they have low wiring costs (i.e., few edges) but allow fast communication between any two nodes in a network [48]. Previous research has suggested that semantic networks are small-world-like [49]. Small-world networks are commonly seen in language (and other domains), perhaps because they emerge from a simple preferential attachment learning mechanism [49, 50], in which newly learned words are more likely to connect to other high-degree words in an existing semantic network than to low-degree words.

Small-worldness can be quantified as the ratio of a network's clustering coefficient relative to random network, over the ratio of a network's average shortest-path length relative to a random network [45]. Networks with a small-world coefficient greater than one are said to be small-world networks. We found that PAD networks are significantly less small-world like compared to NC networks, $M_{NC} = 1.92, M_{PAD} = 1.21, t(107.2) = 4.91, p < .001$, suggesting the efficient interconnectivity of healthy semantic networks is degraded in AD patients.

In addition, PAD patients typically had a higher value for their perseveration parameter p_{emit}, $M_{NC} = .07, M_{PAD} = .34, t(42.3) = 5.34, p < .001$. Under the noisy censored random walk framework, this indicates that the internal monitoring process of PAD patients (deciding whether a word has been said previously) is impaired relative to NC participants. This may be expected, given that PAD patients have higher rates of perseveration in the data, though a higher rate of perseverations in the fluency data does not guarantee a higher value for p_{emit}. The reason for this is that the network structure affects the total number of opportunities for a perseveration to occur. For example, in a fully connected network, an uncensored random walk of a fixed length will produce fewer perseverations than an uncensored random walk of the same length on a linear network.

3.2. Adjusted Network Properties. While we observed many differences between PAD and NC networks, it is difficult to judge whether these differences are due to the mental representations of the two groups or whether they emerge because PAD networks are, on average, generated from a smaller amount of data than NC networks. We adjust for this by constructing corresponding mock networks for each participant network. As described in the "Participant Networks and Mock Networks" subsection, these networks were constructed by permuting the fluency lists of each participant and generating a new network using U-INVITE. We then compute delta metrics by subtracting a participant network's measure from the average of the mock networks. For example,

$$\Delta_{aspl} = G_{aspl} - \left(\frac{\sum_{k=1}^{50} D_{aspl}^{k}}{50} \right) \tag{11}$$

where G_{aspl} denotes the average shortest-path length of participant network G and D_{aspl}^{k} denotes the average shortest-path length of mock network k yoked to participant network

Table 3: Comparison of logistic regression models.

Baseline model (AIC = 52.4)			Maximal model (AIC = 57.3)			Stepwise model (AIC = 45.5)		
Factor	z-value	p-value	Factor	z-value	p-value	Factor	z-value	p-value
Num responses	4.31	< .001*	Num responses	2.39	< .017*	Num responses	3.17	.002*
Perseveration rate	3.71	< .001*	Perseveration rate	1.34	.18	Perseveration rate	1.98	.047*
Education	.77	.44	Education	1.07	.29			
			p_{emit}	2.06	.039*	p_{emit}	2.11	.035*
			Δ Mean degree	1.58	.11	Δ Mean degree	2.21	.027*
			Δ Diameter	1.08	.28	Δ Diameter	1.59	.11
			Diameter	.88	.38			
			Mean degree	1.26	.21			
			Density	.06	.95			
			Shortest-path length	.80	.42			
			Δ Shortest-path length	.73	.47			
			Small-worldness	.66	.51			
			Num nodes	.56	.58			

G. (Here, we use 50 in the denominator because we generate 50 mock networks for each participant's network.)

While both NC and PAD networks have a smaller mean degree compared to their mock counterparts, the difference between the mock and participant networks is smaller for PAD networks than for NC networks, $M_{NC} = -0.81$, $M_{PAD} = -0.26$, $t(105.4) = 6.39$, $p < .001$. The same pattern is true for the networks' median degree, $M_{NC} = -0.38$, $M_{PAD} = -.15$, $t(123) = 2.25$, $p = .026$. Both NC and PAD networks have a larger average shortest-path length compared to mock networks, but again PAD networks are significantly closer to their mock counterparts, $M_{NC} = 0.67$, $M_{PAD} = 0.36$, $t(123) = 3.66$, $p < .001$. NC and PAD networks also have larger diameters than their corresponding mock networks, though PAD networks are closer to their mock networks, $M_{NC} = 2.43$, $M_{PAD} = 1.24$, $t(123) = 2.97$, $p = .004$. Collectively, these results suggest that, in many ways, PAD networks more closely resemble networks generated from randomly generated (i.e., permuted) fluency lists. In contrast, NC networks are quite distinct from networks estimated from randomly generated lists.

While both NC and PAD networks are less dense and less clustered relative to their mock network counterparts, the delta scores themselves do not differ between groups for either density ($p = .47$) or clustering coefficient ($p = .91$).

We do not provide comparisons for small-worldness or p_{emit} adjusted by their mock networks (though their raw values are listed in Table 2). Unlike other network measures, small-worldness is explicitly measured as a ratio relative to the clustering and shortest-path length of a random (Erdös-Rényi) network (see [45]), so no correction is needed. Additionally, p_{emit} is not inherently correlated with network size and does not need to be corrected.

3.3. Relation between Network Measures and Alzheimer's Diagnosis. We used logistic regression to identify associations between network measures and participant diagnosis (NC or PAD) under several different models (interaction terms were excluded to avoid a combinatorial explosion of parameters). In clinical settings, the semantic fluency task is often scored by examining only the total number of responses given and the perseveration rate. In a baseline model, we used these two factors as independent variables along with years of education, which is widely believed to be associated with Alzheimer's disease [51]. Both number of responses and perseveration rate were significantly associated with diagnosis ($p < .001$), as was the model as a whole ($p < .001$, $AIC = 52.38$, null deviance = 158.2, residual deviance = 44.4). Though PAD participants reliably differ from NC participants in years of education, $t(60.9) = 2.56$, $p = .013$, years of education was not significantly associated with diagnosis in the baseline model ($p = .44$) after controlling for other factors in the model.

We compared this baseline model to a maximal model that included ten additional factors. Along with the three factors in the baseline model, we included each of the network measures that are correlated with performance on the Mini-Mental State Exam (i.e., those measures shown in Figure 3: small-worldness, number of nodes, density, mean degree, Δ mean degree, shortest-path length, Δ shortest-path length, p_{emit}, diameter, and Δ diameter.) This maximal model also explained a significant portion of the variance in participant diagnoses ($p < .001$, $AIC = 57.3$, residual deviance = 29.3). See Table 3.

We also conducted an exploratory step-wise regression model using bidirectional elimination starting with the maximal model. The best fit model ($p < .001$, $AIC = 45.5$, residual deviance = 33.5) contained five factors: the total number of responses, perseveration rate, Δ mean degree, p_{emit}, and Δ diameter. Four factors were individually significant ($p < .05$) while one (Δ diameter) was not ($p = .11$). This model outperformed both the baseline and maximal models as measured by AIC, a model selection criterion that penalizes models with more parameters [52].

In addition, we performed a cross-validation of the data to predict the diagnosis of each individual using each of the three models. Cross-validation was performed using split-halves, sampled randomly while preserving the overall ratio of NC and PAD participants in each half (i.e., 67% NC in each training sample). This procedure was repeated on the dataset 5,000 times.

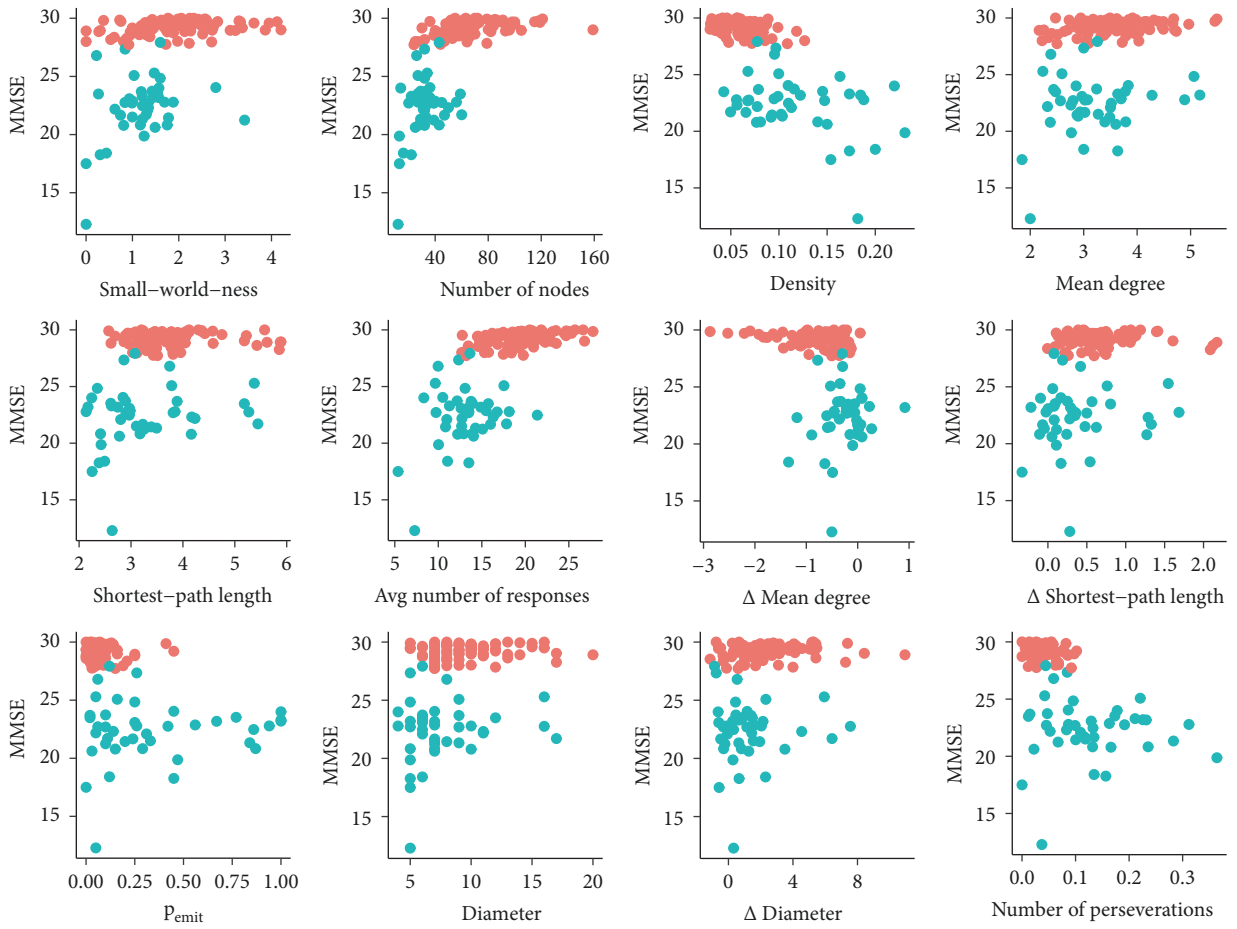


FIGURE 3: Each of the factors we identified as distinguishing between PAD and NC (except Δ median degree, correlation $p = .11$) are plotted with respect to scores on the Mini-Mental State Exam (MMSE) (MMSE scores were unavailable for 7 participant visits out of 1,047). Blue dots indicate patients diagnosed with PAD and red dots indicate those diagnosed as NC. Clinicians had access to patient MMSE scores when making their diagnosis, but did not use semantic fluency data to make their diagnoses. All correlations are significant ($p < .005$, uncorrected for multiple comparisons). Many of these correlations appear to be driven by a restriction in range of the MMSE scores for NC patients. Only the first three factors in the top row (small-worldness, number of nodes, and density) are correlated significantly with MMSE ($p < .05$, uncorrected) when restricted to PAD patients.

TABLE 4: The hits, misses, false alarms, and correct rejections for each model are shown below, averaged across 5,000 split halves.

	Baseline	Stepwise	Maximal
Hits	36.02	36.67	38.11
Misses	5.02	4.37	2.93
False alarms	4.49	4.51	2.00
Correct rejections	80.49	80.47	82.98

Both the step-wise model and the maximal model outperformed the baseline model in predicting diagnoses using measures of accuracy (average $ACC_{baseline} = 92.5\%$, $ACC_{stepwise} = 92.9\%$, $ACC_{maximal} = 96.1\%$) and F1 scores (average $F1_{baseline} = .883$, $F1_{stepwise} = .891$, $F1_{maximal} = .939$). An F1 score denotes the harmonic average of precision and recall. It is used frequently in signal detection analyses to balance the need to correctly predict positive cases and avoid false alarms. It ranges from 0 to 1, where 1 represents perfect precision and recall. A break-down of hits, misses, false alarms, and correct rejections for each model is shown in Table 4. These results suggest that network factors may aid in predicting patient diagnosis. However, because the factors in each model were chosen based on observing the whole dataset, future work is needed to validate these models on an independent dataset.

4. Conclusion

Using a longitudinal corpus of semantic fluency data, we estimated animal semantic networks for individuals categorized as either healthy (normal control) or probable Alzheimer's patients. These networks revealed systematic differences between the mental representations of the two groups. Healthy semantic networks were larger, less dense, and contained a higher number of associations per concept (i.e., higher mean degree). Using a bootstrap approach to generate mock networks, we found that Alzheimer's networks are significantly closer to mock networks generated from random permutations of the participants' data. In contrast, healthy networks are more small-world-like, consistent with prior literature on semantic networks [49] and human language [53], and drawing a parallel to a finding that large-scale brain networks in AD patients are also less small-world-like than healthy controls [54].

These results corroborate previous results that have shown atypical associations in the semantic representations of AD patients [46, 55]. Using current best practices for estimating semantic networks, our results corroborate those of Lerner et al. [12] finding that unadjusted AD semantic networks are less dense and less small-world-like, having smaller diameter and higher average node degree.

However those differences between AD and control semantic networks should be interpreted in light of the processes and data used to construct them: using the censored random walk model (or a naïve random walk model used by [12]), smaller data sets (i.e., fewer fluency lists or shorter lists) that are evenly generated at random produce smaller networks (i.e., fewer nodes), which distorts many network properties. In contrast to previous work, we adjusted for this potential confound by comparing each network to a null model (i.e., mock networks) that assumes items in a fluency list do not have any sequential dependencies. This procedure revealed that some findings, such as the difference in the density of networks between groups, may be artifacts of the methodology used to construct networks. Future work should test the robustness of our results against different methods for constructing networks (e.g., [30]) as well as other null models: for example, Chan et al. [5] construct networks using triadic comparison data so that all participant networks have the same number of nodes, while Kenett et al. [56] compare estimated networks to Erdös-Rényi random networks of the same size. The field has yet to come to a consensus on the most appropriate null model to use for comparing networks, though it is likely that each of these approaches have strengths and weaknesses.

Previous attempts at mapping semantic memory in patients with semantic impairments have been criticized as being methodologically inadequate [15]. Part of this criticism stems from constructing group networks by averaging across the representations of individuals, who likely have unique impairments. In contrast to previous research [5, 12], our study is the first to estimate semantic network representations of individual patients with AD. In addition, Verheyen et al. [16] suggest that previous methods used to map semantic representations (such as multidimensional scaling or singular value decomposition; see [57]) do not produce stable representations, even when generated from random samples from the same individual's data. While this is a concern, our method of generating semantic networks is both theoretically and mathematically distinct from these criticized approaches. The work they criticized used estimation techniques that were exchangeable, which means that the item order in a list did not affect the estimated representation. Our method is nonexchangeable. Changing the order of items in a list affects the probability of estimated networks due to earlier items being much less likely to be affected by censoring than later items. Further, Zemla and Austerweil [31] found that the deterministic censoring version of our method for estimating networks from fluency data was empirically valid: estimated edges were judged to have high semantic similarity in pairwise similarity ratings compared to nonedges.

Similarly, Verheyen et al. [16] suggest that perhaps semantic fluency data cannot reliably estimate semantic representations because the semantic fluency task taps into other cognitive processes in addition to representation. We agree, and we caution using our results to advocate for a purely storage deficit (as opposed to retrieval deficit) in AD. The censored random walk model of memory retrieval on which our network inference method is based [31] assumes that semantic retrieval is biased towards items that are semantically similar to recently retrieve items. While this is generally accepted to be true, mental search may also rely on frontal lobe processes that are independent of the semantic representation [58]. Future work may modify the censored random walk model that allow for random or strategic jumps [33, 41] that reflect an explicit cluster switching component (or "restarts" of the search process) and possibly distinguish the influence of representation versus executive functioning on mental search. Though previous work suggests these jumps are not necessary to model healthy fluency behavior, they may play a role when modeling behavior from populations with memory disorders. Furthermore, we find differences in a working memory monitoring process of NC and AD patients (as evidenced by differences in p_{emit}) that more closely align with retrieval rather than storage deficits. It is likely that the semantic impairments of AD patients are due in part to both storage and retrieval deficits, and our results suggest one model that explains these impairments through an interaction of the two.

One limitation of our current approach is that we only look at the structure of semantic networks in the animal category. Although the animal fluency task is extremely common in the psychology literature and in clinical practice, it may not be the best proxy for an individual's semantic memory as a whole. Chan, Salmon, and De La Pena [59] found that while semantic representations of the animal category are impaired in AD, the tools category remained largely intact. Though a focus on the animal category may be useful for identifying semantic decline in AD patients, analysis of a broader spectrum of categories could provide a more wholistic view of semantic memory impairment in AD.

Finally, we found that the inclusion of network measures in a logistic regression model improved prediction of participant diagnosis, even after adjusting for the increased number

of parameters. This suggests that a network-based approach may explain more variance than traditional approaches to scoring the semantic fluency task and could improve identification of patients with Alzheimer's disease.

The current research highlights potential differences in the structural properties of semantic networks of individuals with and without Alzheimer's disease, yet raises additional questions about the computational processes that might produce these changes. As patients transition from healthy to impaired, are changes to their semantic network better modeled by the addition or removal of edges, or some combination of both? Are edged removed (or added) at random in the network, or do these changes occur at predictable locations in the network? For example, are edge changes more likely at high-degree nodes? Do they spread from "infected" nodes? Our current results suggest that a process that adds spurious edges at random might be a good candidate for exploration, but further research is needed.

In the above analyses, we consider only two diagnostic points: healthy (NC) and probable Alzheimer's diagnoses (PAD). Future research should examine individuals with Mild Cognitive Impairment, an intermediary phase between healthy and Alzheimer's disease, as well as track the networks of individuals as they evolve over time. In doing so, it may be possible to uncover the dynamic processes that explain the transition between healthy and impaired networks. While biological models of how AD spreads have been postulated (e.g., [60]), no such processes have been proposed on the algorithmic level for semantic network degradation.

We believe our results represent the first attempt to estimate individual semantic networks from a psychologically plausible process model in order to assess memory impairment. Future research can extend this approach in many ways. Many clinical populations other than Alzheimer's patients are impaired on the semantic fluency task—including those with Huntington's disease [61], frontotemporal dementia [62], and semantic dementia [63]. Studies have found that these groups may have distinct behavioral profiles on the semantic fluency task, and perhaps their semantic networks are distinct as well.

Overall, we find that a network-based analysis of semantic fluency data may improve diagnosticity of Alzheimer's disease, while providing clues to the cognitive mechanisms that lead to impairment on the semantic fluency task. This approach may provide a useful tool for assessing other neuropsychiatric disorders and provide new insight into how we store and retrieve semantic knowledge.

Data Availability

The data used to support the findings of this study are included within the Supplementary Materials.

Conflicts of Interest

The authors have no conflicts of interest to declare, financial or otherwise.

Acknowledgments

We thank Bill Heindel, David Salmon, and the University of California-San Diego Shiley-Marcos Alzheimer's Disease Research Center for providing the longitudinal fluency data set and technical assistance. We would also like to thank Elizabeth Pettit, Jacqueline Erens, and Jacob Hurlburt for help with data transcription. This research was performed using the compute resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. Support for this research was provided by NIH R21AG0534676 and the Office of the VCGRE at UW-Madison with funding from the WARF. The fluency dataset was collected by the University of California-San Diego Shiley-Marcos Alzheimer's Disease Research Center with support by NIH AG05131.

Supplementary Materials

Data and analysis code for this manuscript are available at https://osf.io/j6qea/. Included in the these materials are the raw fluency data used for analysis (ucsd_fluency_for_snafu_20180518.csv), networks estimated from fluency data using U-INVITE (networks.zip), network statistics computed from these networks (ad_graphs_usf_persev.json), a matrix of correlations between factors (correlation_matrix.pdf), and R analysis code (analysis.r). *(Supplementary Materials)*

References

[1] T. Vos, C. Allen, M. Arora, M. Barber R, A. Bhutta Z, and A. Brown, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015," *The Lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.

[2] F. J. Huff, S. Corkin, and J. H. Growdon, "Semantic impairment and anomia in Alzheimer's disease," *Brain and Language*, vol. 28, no. 2, pp. 235–249, 1986.

[3] D. Howard and K. E. Patterson, *The Pyramids and Palm Trees Test: A Test of Semantic Access from Words and Pictures*, Thames Valley Test Company, 1992.

[4] J. R. Hodges and K. Patterson, "Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications," *Neuropsychologia*, vol. 33, no. 4, pp. 441–459, 1995.

[5] A. S. Chan, N. Butters, D. P. Salmon, S. A. Johnson, J. S. Paulsen, and M. R. Swenson, "Comparison of the semantic networks in patients with dementia and amnesia," *Neuropsychology*, vol. 9, no. 2, pp. 177–186, 1995.

[6] C. Randolph, A. R. Braun, T. E. Goldberg, and T. N. Chase, "Semantic fluency in alzheimer's, parkinson's, and huntington's disease: dissociation of storage and retrieval failures," *Neuropsychology*, vol. 7, no. 1, pp. 82–88, 1993.

[7] D. P. Salmon, N. Butters, and A. S. Chan, "The deterioration of semantic memory in Alzheimer's disease," *Canadian Journal of Experimental Psychology*, vol. 53, no. 1, pp. 108–116, 1999.

[8] R. D. Nebes, D. C. Martin, and L. C. Horn, "Sparing of semantic memory in Alzheimer's disease," *Journal of Abnormal Psychology*, vol. 93, no. 3, pp. 321–330, 1984.

[9] R. D. Nebes, "Semantic memory in alzheimer's disease," *Psychological Bulletin*, vol. 106, no. 3, pp. 377–394, 1989.

[10] D. S. Roy, A. Arons, T. I. Mitchell, M. Pignatelli, T. J. Ryan, and S. Tonegawa, "Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease," *Nature*, vol. 531, no. 7595, pp. 508–512, 2016.

[11] M. N. Jones, J. Willits, S. Dennis, and M. Jones, "Models of Semantic Memory Models of Semantic Memory," *Oxford Handbook of Mathematical and Computational Psychology*, pp. 232–254, 2015.

[12] A. J. Lerner, P. K. Ogrocki, and P. J. Thomas, "Network graph analysis of category fluency testing," *Cognitive and Behavioral Neurology*, vol. 22, no. 1, pp. 45–52, 2009.

[13] D. Caine and J. R. Hodges, "Heterogeneity of semantic and visuospatial deficits in early Alzheimer's disease," *Neuropsychology*, vol. 15, no. 2, pp. 155–164, 2001.

[14] J. L. Cummings, "Cognitive and behavioral heterogeneity in Alzheimer's disease: Seeking the neurobiological basis," *Neurobiology of Aging*, vol. 21, no. 6, pp. 845–861, 2000.

[15] G. Storms, T. Dirikx, J. Saerens, S. Verstraeten, and P. P. De Deyn, "On the use of scaling and clustering in the study of semantic deficits," *Neuropsychology*, vol. 17, no. 2, pp. 289–301, 2003.

[16] S. Verheyen, W. Voorspoels, J. Longenecker, D. R. Weinberger, B. Elvevåg, and G. Storms, "Invalid assumptions in clustering analyses of category fluency data: Reply to Sung, Gordon and Schretlen (2015)," *Cortex*, vol. 75, pp. 255–259, 2016.

[17] A. M. Collins and M. R. Quillian, "Retrieval time from semantic memory," *Journal of Verbal Learning and Verbal Behavior*, vol. 8, no. 2, pp. 240–247, 1969.

[18] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing," *Psychological Review*, vol. 82, no. 6, pp. 407–428, 1975.

[19] D. J. Watts, "The "new" science of networks," *Annual Review of Sociology*, vol. 30, pp. 243–270, 2004.

[20] Y. N. Kenett, D. Anaki, and M. Faust, "Investigating the structure of semantic networks in low and high creative persons," *Frontiers in Human Neuroscience*, vol. 8, pp. 1–16, 2014.

[21] K. Borodkin, Y. N. Kenett, M. Faust, and N. Mashal, "When pumpkin is closer to onion than to squash: The structure of the second language lexicon," *Cognition*, vol. 156, pp. 60–70, 2016.

[22] D. U. Wulff, T. T. Hills, and M. R. Lachman, "The aging lexicon: Differences in the semantic networks of younger and older adults," in *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pp. 907–912, Cognitive Science Society, Austin, TX, USA, 2016.

[23] A. Benton, K. D. S. Hamsher, and A. Sivan, *Multilingual Aphasia Examination*, AJA Associates, Iowa City, IA, USA, 1994.

[24] N. Helm-Estabrooks, *Cognitive linguistic quick test: CLQT*, PsychCorp, 2001.

[25] S. Weintraub, D. Salmon, N. Mercaldo et al., "The Alzheimer's Disease Centers' Uniform Data Set (UDS): The neuropsychologic test battery," *Alzheimer Disease & Associated Disorders*, vol. 23, no. 2, pp. 91–101, 2009.

[26] A. K. Troyer, M. Moscovitch, G. Winocur, L. Leach, and M. Freedman, "Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease," *Journal of the International Neuropsychological Society*, vol. 4, no. 2, pp. 137–143, 1998.

[27] S. Pekkala, M. L. Albert, A. Spiro III, and T. Erkinjuntti, "Perseveration in Alzheimer's disease," *Dementia and Geriatric Cognitive Disorders*, vol. 25, no. 2, pp. 109–114, 2008.

[28] S. V. S. Pakhomov, L. E. Eberly, and D. S. Knopman, "Recurrent perseverations on semantic verbal fluency tasks as an early marker of cognitive impairment," *Journal of Clinical and Experimental Neuropsychology*, vol. 40, no. 8, pp. 832–840, 2018.

[29] K. D. Mueller, S. L. Allison, R. L. Koscik, E. Jonaitis, B. T. Christian, and T. J. Betthauser, "Verbal fluency measures are associated with alzheimer's disease biomarkers in clinically unimpaired late middle-aged adults from the wisconsin registry for alzheimer's prevention," *Alzheimer's & Dementia*, vol. 14, no. 7, pp. P23–P24, 2018.

[30] J. Goñi, G. Arrondo, J. Sepulcre et al., "The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory," *Cognitive Processing*, vol. 12, no. 2, pp. 183–196, 2011.

[31] J. C. Zemla and J. L. Austerweil, "Estimating semantic networks of groups and individuals from fluency data Estimating semantic networks of groups and individuals from fluency data," *Computational Brain and Behavior*, vol. 1, no. 1, pp. 36–58, 2018.

[32] A. K. Troyer, M. Moscovitch, and G. Winocur, "Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults," *Neuropsychology*, vol. 11, no. 1, pp. 138–146, 1997.

[33] J. T. Abbott, J. L. Austerweil, and T. L. Griffiths, "Random walks on semantic networks can resemble optimal foraging," *Psychological Review*, vol. 122, no. 3, pp. 558–569, 2015.

[34] M. Miozzo, S. Fischer-Baum, and E. Caccappolo-van Vliet, "Perseverations in Alzheimer's disease: memory slips?" *Cortex*, vol. 49, no. 8, pp. 2028–2039, 2013.

[35] D. P. Salmon, W. C. Heindel, and K. L. Lange, "Differential decline in word generation from phonemic and semantic categories during the course of Alzheimer's disease: Implications for the integrity of semantic memory," *Journal of the International Neuropsychological Society*, vol. 5, no. 7, pp. 692–703, 1999.

[36] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of alzheimer's disease report of the nincds-adrda work group under the auspices of department of health and human services task force on alzheimer's disease," *Neurology*, vol. 34, no. 7, pp. 939–944, 1984.

[37] M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""Mini mental state": A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

[38] D. Galasko, L. A. Hansen, R. Katzman et al., "Clinical-neuropathological correlations in Alzheimer's disease and related dementias," *JAMA Neurology*, vol. 51, no. 9, pp. 888–895, 1994.

[39] M. S. Albert, S. T. DeKosky, D. Dickson et al., "The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 270–279, 2011.

[40] K. S. Jun, X. Zhu, T. T. Rogers, Z. Yang et al., "Human memory search as initial-visit emitting random walk," in *Advances in Neural Information Processing Systems*, pp. 1072–1080, 2015.

[41] J. C. Zemla and J. L. Austerweil, "Modeling semantic fluency data as search on a semantic," in *Proceedings of the 38th annual meeting of the cognitive science society*, pp. 3646–3651, 2017.

[42] P. G. Doyle and J. L. Snell, *Random Walks And Electric Networks*, Mathematical Association of America, 1984.

[43] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber, "The University of South Florida free association, rhyme, and word fragment norms," *Behavior Research Methods, Instruments, and Computers*, vol. 36, no. 3, pp. 402–407, 2004.

[44] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 1–54, 2002.

[45] M. D. Humphries and K. Gurney, "Network 'small-world-ness': a quantitative method for determining canonical network equivalence," *PLoS ONE*, vol. 3, no. 4, Article ID e0002051, 2008.

[46] A. S. Chan, N. Butters, D. P. Salmon, and K. A. McGuire, "Dimensionality and clustering in the semantic network of patients with alzheimer's disease," *Psychology and Aging*, vol. 8, no. 3, pp. 411–419, 1993.

[47] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[48] D. S. Bassett and E. T. Bullmore, "Small-world brain networks," *The Neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.

[49] M. Steyvers and J. B. Tenenbaum, "The large-scale structure of semantic networks: statistical analyses and a model of semantic growth," *Cognitive Science*, vol. 29, no. 1, pp. 41–78, 2005.

[50] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[51] A. Ott, M. M. B. Breteler, F. van Harskamp et al., "Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study," *British Medical Journal*, vol. 310, no. 6985, pp. 970–973, 1995.

[52] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, Eds., pp. 267–281, Academiai Kiado, Budapest, Hungary, 1973.

[53] R. F. I. Cancho and R. V. Solé, "The small world of human language," *Proceedings of the Royal Society B Biological Science*, vol. 268, no. 1482, pp. 2261–2265, 2001.

[54] X. Zhao, Y. Liu, X. Wang et al., "Disrupted small-world brain networks in moderate Alzheimer's disease: a resting-state fMRI study," *PLoS ONE*, vol. 7, no. 3, Article ID e33540, 2012.

[55] A. S. Chan, N. Butters, J. S. Paulsen, D. P. Salmon, M. R. Swenson, and L. T. Maloney, "An assessment of the semantic network in patients with Alzheimer's disease," *Cognitive Neuroscience*, vol. 5, no. 2, pp. 254–261, 1993.

[56] Y. N. Kenett, D. Wechsler-Kashi, D. Y. Kenett, R. G. Schwartz, E. Ben-Jacob, and M. Faust, "Semantic organization in children with cochlear implants: Computational analysis of verbal fluency," *Frontiers in Psychology*, vol. 4, pp. 1–11, 2013.

[57] K. Sung, B. Gordon, and D. J. Schretlen, "Semantic structure can be inferred from category fluency tasks via clustering analyses: Reply to Voorspoels et al. (2014)," *Cortex*, vol. 75, pp. 249–254, 2016.

[58] R. M. Birn, L. Kenworthy, L. Case et al., "Neural systems supporting lexical search guided by letter and semantic category cues: A self-paced overt response fMRI study of verbal fluency," *NeuroImage*, vol. 49, no. 1, pp. 1099–1107, 2010.

[59] A. S. Chan, D. P. Salmon, and J. De La Pena, "Abnormal semantic network for "animals" but not "tools" in patients with Alzheimer's disease," *Cortex*, vol. 37, no. 2, pp. 197–217, 2001.

[60] A. Raj, A. Kuceyeski, and M. Weiner, "A network diffusion model of disease progression in dementia," *Neuron*, vol. 73, no. 6, pp. 1204–1215, 2012.

[61] J. D. Henry, J. R. Crawford, and L. H. Phillips, "A meta-analytic review of verbal fluency deficits in Huntington's disease," *Neuropsychology*, vol. 19, no. 2, pp. 243–252, 2005.

[62] K. Rascovsky, D. P. Salmon, L. A. Hansen, L. J. Thal, and D. Galasko, "Disparate letter and semantic category fluency deficits in autopsy-confirmed frontotemporal dementia and Alzheimer's disease," *Neuropsychology*, vol. 21, no. 1, pp. 20–30, 2007.

[63] A.-L. R. Adlam, K. Patterson, T. T. Rogers et al., "Semantic dementia and fluent primary progressive aphasia: Two sides of the same coin?" *Brain*, vol. 129, no. 11, pp. 3066–3080, 2006.

Hindawi
Complexity
Volume 2019, Article ID 1918504, 15 pages
https://doi.org/10.1155/2019/1918504

Research Article

Mediation Centrality in Adversarial Policy Networks

Stefan M. Herzog[1] **and Thomas T. Hills**[2]

[1]*Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany*
[2]*Department of Psychology, University of Warwick, UK*

Correspondence should be addressed to Stefan M. Herzog; herzog@mpib-berlin.mpg.de

Received 7 September 2018; Revised 21 December 2018; Accepted 10 January 2019; Published 30 April 2019

Guest Editor: Yoed Kenett

Conflict resolution often involves mediators who understand the issues central to both sides of an argument. Mediators in complex networks represent key nodes that are connected to other key nodes in opposing subgraphs. Here we introduce a new metric, *mediation centrality*, for identifying good mediators in adversarial policy networks, such as the connections between individuals and their reasons for and against the support of controversial topics (e.g., state-financed abortion). Using a process-based account of reason mediation we construct bipartite adversarial policy networks and show how mediation defined over subgraph projections constrained to reasons representing opposing sides can be used to produce a measure of mediation centrality that is superior to centrality computed on the full network. We then empirically illustrate and test mediation centrality in a "policy fluency task," where participants generated reasons for or against eight controversial policy issues (state-subsidized abortion, bank bailouts, forced CO_2 reduction, cannabis legalization, shortened naturalization, surrogate motherhood legalization, public smoking ban, and euthanasia legalization). We discuss how mediation centrality can be extended to adversarial policy networks with more than two positions and to other centrality measures.

1. Introduction

Adversarial systems can be defined as systems composed of individuals with opposing views, such as Democrats versus Republicans in US politics or Leave versus Remainers in the Brexit discussion. Numerous recent studies have investigated the development of adversarial information environments that can isolate individuals from the views of their opponents, such as echo chambers and filter bubbles [1–3]. This isolation can lead to overconfidence and further polarization and, counter-intuitively, may be especially prominent in information-rich environments [4]. These systems often form over ideological divisions and extend even to the truth value of science, with the end result that such groups rarely see eye-to-eye and are severely insulated from one another in relation to beliefs and social contacts (e.g., [5, 6]).

To make progress on controversial issues in adversarial systems, it can be useful to identify individuals who are best able to help collective problem solving. Such individuals should be able to guide others towards recognizing and acknowledging the beliefs and values of individuals on different sides of an issue [7]. For example, a capacity for perspective taking—the ability to understand and acknowledge views on different sides of an issue—is one of the most effective tools of a good negotiator [8]. Similarly, convergent framing that identifies a collectively recognized description of the problem can help facilitate conflict resolution [9]. From the perspective of conflict resolution [10], it is this ability to recognize the collective perspectives that is a defining characteristic of a good *mediator* [11]. Such good mediators maximize the opportunity that the majority of individuals on each side of the issue can agree on what the disagreement really comes down to.

Adversarial systems of the kind described above can be considered complex networks, where individuals are connected to other individuals by acknowledgement of shared reasons supporting opposing sides of an issue. Although network science has proposed many centrality metrics for identifying key nodes in a variety of contexts [12–16], we know of no metric for identifying mediators in adversarial

systems, and more specifically adversarial policy networks, where a network is adversarial because it contains information representing more than one position about the policy. Adversarial policy networks can be represented by bipartite networks, where individuals are connected by edges to the reasons they acknowledge. Figure 1(a) provides an empirical example of such a bipartite adversarial policy network based on reasons—and the individuals who recognize those reasons—concerning the policy issue of reducing the minimum number of years of residence to become a naturalized citizen in Switzerland (based on empirical results of a study described later in the paper). In such adversarial policy networks, the reasons can support only one of several sides of an issue. Such a network can be projected onto persons (where individuals are connected if they share at least one reason) or reasons (where two reasons are connected if they are produced by the same person; Figure 1(b)). By constraining the reasons to be on one side of the issue one can further describe subgraphs of individuals who are connected in relation to either pro or con reasons (Figure 1(c)).

In bipartite adversarial policy networks, there are reasons for and against the policy as well as individuals who acknowledge different subsets of those reasons, with some individuals recognizing reasons on both sides of the issue. A good mediator in this space is someone with high centrality in all subgraphs. Importantly, by this definition, a good mediator is not necessarily someone who recognizes the most reasons or the reasons that would make the most people happy, nor is it someone who acknowledges an equal amount of reasons among the various positions or even a person who recognizes those reasons that would best cover the reason space as defined by what people collectively acknowledge (a metric, representativeness, which we describe below). Each of these attributes can be gamed simply by adding more people or poor quality reasons. The method we describe below is immune to such subterfuge.

The central contribution of this paper is the introduction and investigation of *mediation centrality*, a network measure for identifying mediators in bipartite adversarial policy networks. Mediation centrality is computed by combining centrality metrics from subgraph projections where the projections are defined in relation to different sets of reasons.

In what follows, we first define bipartite adversarial policy networks and then describe our novel mediation metric for identifying graph mediators on these networks. We then evaluate this mediation metric using simulated bipartite adversarial policy networks and show that mediation centrality captures the notion of a good mediator who is the best-recognizer-of-best-recognized-reasons in a hypothetical discussion that follows an associative path through the argument space. A good mediator in this space would have the most to contribute to this discussion. We then empirically illustrate mediation centrality in a "policy fluency task," where participants generated reasons for or against a range of controversial policy issues.

2. Mediation Centrality

2.1. Bipartite Adversarial Policy Networks, Mediation, and Centrality Metrics. Bipartite adversarial policy networks can be represented as graphs, $G(V, E)$, where vertices, V, are composed of individuals, I, and reasons, R, with edges, E, connecting individuals to reasons (Figure 1(a)). For the network to be adversarial, reasons represent positions with respect to the policy and are therefore exclusive to one subgraph. As we note below, this can be extended to any number of positions, but for ease of exposition we assume that reasons can only be either *for* or *against* the policy.

The projection of reasons onto individuals gives a graph $G_I(I, E)$ where individuals i and j are connected in the resulting adjacency matrix if they share at least one reason, $k \in R$, (Figure 1(b), left graph), such that

$$\mathbf{A}_{ij} = \sum_k R_{i,k} R_{j,k} \tag{1}$$

where $R_{i,k}$ has a value of 1 if reason k is held by individual i and 0 otherwise. Similarly, one can form projections onto reasons (Figure 1(b), right graph).

The different sides of the position can be represented by $G_+(I, R_+; E)$ and $G_-(I, R_-; E)$, representing the subgraphs formed by constraining reasons to those either for or against the issue, respectively. Forming the projections onto individuals as above, we get $G_{I,+}$ and $G_{I,-}$, respectively, which represent individuals' connectivity solely driven by either pro or con reasons, respectively (Figure 1(c)).

There are a variety of centrality metrics that could be computed on each of the subgraph projections, such as degree centrality, betweenness centrality, and closeness centrality. Mediation centrality can be generalized to each of these metrics as we will discuss below. However, because we are considering mediation in the context of a domain where reasons are represented in individuals' minds, we are interested in how ideas are connected between people. In particular, we are interested in the process of a hypothetical fruitful discussion, where the discussion tracks the structural information defined by associations between people and the reasons they collectively acknowledge. In such a setting, a good mediator is someone who would contribute maximally to this hypothetical discussion because she knows and can introduce the collectively important reasons into the discussion. She therefore is a best-recognizer-of-best-recognized reasons (as we show below). We identify this mediator by making two assumptions: that thought is associative and that people are connected, among other things, by shared ideas.

In "Trayne of Thoughts," Hobbes [17] recognized what has become a truism in contemporary cognitive science: one thought gives rise to another in relation to the association between them [18, 19]. Extending this idea to a collection of individuals in an adversarial policy network, we imagine a simple model of a social process whereby individuals activate one another by their shared reasons, with one individual stating one reason and another responding to that reason with another associated reason that comes to mind. This process can be formally described as a random walk through policy space, where transitions between individuals occur

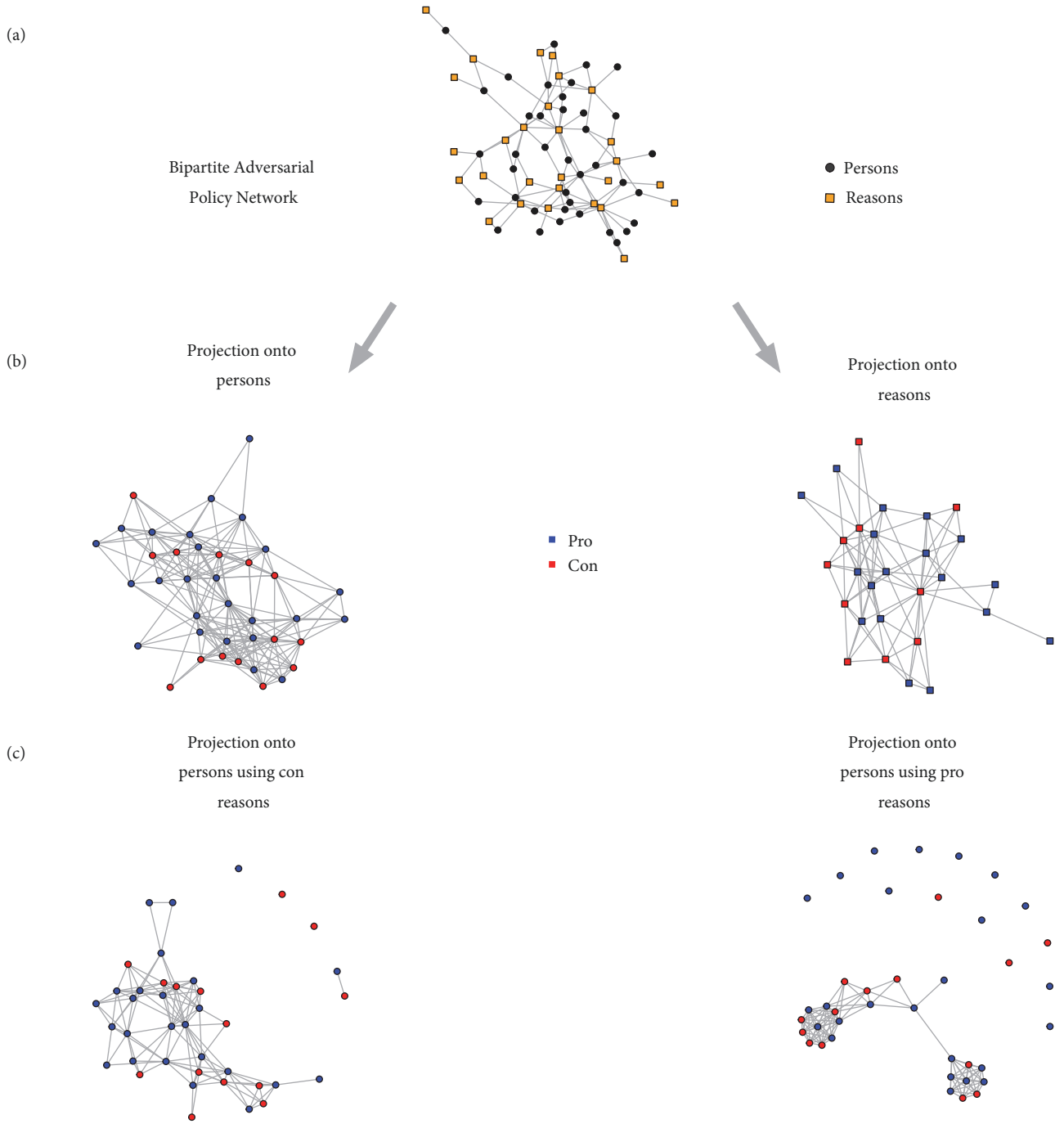


Figure 1: An example of a bipartite adversarial policy network around the issue of speeding up naturalization of foreigners in Switzerland (based on empirical results of a study described later in the paper). (a) The bipartite network composed of persons and the reasons they acknowledge. (b) Projections of the bipartite network onto persons, where individuals are connected if they share at least one reason, or onto reasons, where reasons are connected if they are both produced by the same person. Blue indicates reasons or persons favoring reduced naturalization times and red indicates reasons or persons against reduced naturalization times. (c) The person projections based on bipartite networks constrained to only contain either pro or con reasons. Blue indicates persons favoring reduced naturalization times and red indicates persons against reduced naturalization times; persons' attitudes towards the policy issue were assessed using a separate survey item.

by choosing edges at random in the collective associative representation.

A projection of the bipartite adversarial policy network onto individuals captures this process, whereby individuals are connected if they can activate one another through a shared idea. A random walk over this subspace is equivalent to the process described above. The probability of moving between two nodes in this subspace is described by the

transition matrix **T** and describes a Markov process which converges to a stationary distribution over successive transitions [20].

This stationary distribution can be represented by the vector **x**. Stationarity implies that further transitions do not affect the distribution, such that

$$\mathbf{x} = \mathbf{T}\mathbf{x}, \tag{2}$$

where **x** is the eigenvector associated with the largest eigenvalue of **T**. **T** is the normalized adjacency matrix represented as follows:

$$\mathbf{T}_{ij} = \frac{\mathbf{A}_{ij}}{\sum_{k=1}^{N} \mathbf{A}_{kj}} \tag{3}$$

where N is the total number of individuals. The values of the stationary distribution **x** correspond to a special case of *PageRank* [21] for undirected graphs. PageRank is a network measure, which has been applied to numerous cognitive and social phenomena (e.g., [19, 22–24]). Roughly speaking, the PageRank of a node corresponds to the probability of finding a random walker at that node, where the walker is subject to a Markov process constrained by the adjacency matrix. Although mediation centrality can be generalized, in principle, to any centrality metric (such as degree centrality, betweenness centrality, and closeness centrality), in the description of mediation centrality that follows we will restrict our investigation to PageRank because it follows the logic outlined above of a discussion constrained by the structure of associative relations between reasons as they occur among people. However, in the discussion we will argue that which centrality metric is most appropriate for any domain will depend on the processes involved in that domain.

2.2. Computing Mediation Centrality over Subgraphs. Centrality measures are routinely computed on the full network and may therefore be considered global centrality measures. As we show later, a global measure of centrality does not capture the notion of a good mediator because mediators need to be able to mediate discussions between different positions. That is, in the context of adversarial policy networks, an individual with high centrality on $G_I(I, E)$ may not be a good mediator across opposing subgraphs. Specifically, they may not acknowledge issues that would make them central to G_{I-} and G_{I+} at the same time. To handle this problem, we define a node's mediation centrality, M, as the harmonic mean of its centrality values across subgraphs

$$M = \frac{2}{1/\mathbf{x}_{i,+} + 1/\mathbf{x}_{i,-}} = \frac{2\mathbf{x}_{i,+}\mathbf{x}_{i,-}}{\mathbf{x}_{i,+} + \mathbf{x}_{i,-}} \tag{4}$$

where $\mathbf{x}_{i,+}$ and $\mathbf{x}_{i,-}$ represent the centrality computed for node i from the subgraphs G_{I+} and G_{I-}, respectively. The harmonic mean captures our intended notion of mediation centrality because it is dominated by the smallest (minimum) centrality across the subspaces. In particular, the right-most part of equation (4) highlights that if either $\mathbf{x}_{i,+}$ or $\mathbf{x}_{i,-}$ is zero, $M = 0$—irrespective of the value for the other centrality (unless it is also zero, then M is undefined).

More generally, the harmonic mean H is a Schur-concave function, which implies that for any positive set of n inputs we have $\min(x_1 \dots x_n) \leq H(x_1 \dots x_n) \leq n \min(x_1 \dots x_n)$. This means that H cannot be made arbitrarily large without also changing the value of its smallest input. In particular, if any input is zero, $H = 0$, irrespective of the values of all other inputs (unless all inputs are zero, then H is undefined). The geometric mean is also a Schur-concave function. However, we chose the harmonic mean in analogy to computations of average speed: when a vehicle travels at rate a and then at rate b for equal distances, then the average rate is the harmonic mean of a and b. Loosely speaking, the average "rate" of a mediator's contribution within a subspace is how often they contribute to the random walk in the associated subgraph (i.e., how often an individual is visited by the random walk). If we assume an analogous concept of equidistant paths through the pro and con argument spaces, the average rate of an individual's contribution is proportional to the harmonic mean of their contributions over subgraphs.

For adversarial policy networks consisting of n subspaces representing n positions, mediation centrality is defined as

$$M = \frac{n}{\sum_{i=1}^{n} \mathbf{x}_i^{-1}}, \tag{5}$$

where $\mathbf{x_i}$ is the centrality for the ith subspace computed from $G_{I,i}$. This conveniently reduces to the standard centrality measure for the case of only one subspace ($n = 1$).

Note that mediation centrality is different from the subgraph centrality described by Estrada & Rodriguez-Velazquez [25], which "counts the times that a node takes part in the different connected subgraphs of the network," such as triangles, four cycles, and so on.

2.3. Mediation Centrality versus Representativeness. Earlier we argued for a process-based measure of mediation that can capture structural relations between social and cognitive processes. We then identified PageRank as a suitable network metric for such a mediation measure. Here we introduce and formalize the notion of *representativeness*, a cognitive measure designed to capture reason coverage in a population. We note that representativeness—unlike mediation centrality—does not incorporate network structure. Comparing mediation centrality to representativeness helps to highlight the potential weaknesses of simple counting measures.

To develop the notion of representativeness, let us assume a unit called a *person–reason*, which represents a reason held by one person. Two person–reasons can reflect two different people who each acknowledge one reason—which may or may not be the same reason—or one individual who acknowledges two different reasons. By this unit, ten people who all acknowledge the same, one reason (= 10 person–reasons) have a less formidable reason space than ten people who acknowledge five reasons each (whether or not they are shared; = 50 person–reasons). If we assume that a reason in an individual's mind represents a slot in the adversarial space, then the total adversarial space for one side of an issue is the sum of the slots, that is, the sum

of all person–reasons. This assumes that reason slots are interchangeable. This may not always be the case as some reasons may be more convincing than others even though they are held by fewer individuals. Incorporating a reason's normative weights is beyond the scope of the current paper (but see [26], for some ideas to build on).

For illustration, Figure 2 depicts an adversarial space consisting of 20 unique reasons and 28 person–reasons with one reason held by 4 individuals, two reasons held by three individuals each, and so on. The extent to which an individual covers this space is a measure of their representativeness ρ. We can therefore define the representativeness of an individual as the sum of the slots (person–reasons) they cover in the adversarial space

$$\rho = \sum_{i \in C} w_i, \tag{6}$$

where for each reason i in that individual's set of reasons, C, we sum the number of individuals w_i who acknowledge that reason. An individual with a higher ρ is an individual who better covers the adversarial space over which C is represented. Accordingly, we can compute ρ_p and ρ_c to indicate the representativeness within the pro and con reason spaces, respectively.

As computed here, a node's representativeness within a subgraph is equivalent to the node's weighted degree in a network, where edge weights reflect the number of shared reasons, plus the node's unweighted degree, representing the number of reasons held by the individual. The addition of the node's unweighted degree could be removed to avoid situations where an individual creates unique, idiosyncratic reasons to amplify their own representativeness. In the present case, we leave this in with the assumption that the individuals in the network reflect the population from which they are sampled. This also fits the framing in the empirical study below where we asked individuals to generate reasons they believe would be held by others.

We define *mediation representativeness* over multiple spaces, $\widehat{\rho}$, by the harmonic mean of the representativeness over subgraphs,

$$\widehat{\rho} = \frac{n}{\sum_{i=1}^{n} \rho_i^{-1}} \tag{7}$$

where n is the number of subspaces. As in the definition of median centrality (see earlier), the harmonic mean best captures our intended meaning of mediation representativeness because it is dominated by the smallest representativeness across the subspaces. This prevents individuals from becoming more representative by merely capturing a larger share of an already well-represented subspace.

Though M and $\widehat{\rho}$ may often be correlated in practice—and indeed are well correlated in the empirical study we describe below—they need not be correlated. To see why, consider a reason network where two candidate mediators have identical Ms and $\widehat{\rho}$s. The first candidate can improve her $\widehat{\rho}$ by listing one more reason on each side of the policy issue. However, her M remains unchanged, as it crucially depends on the reasons being recognized by others.

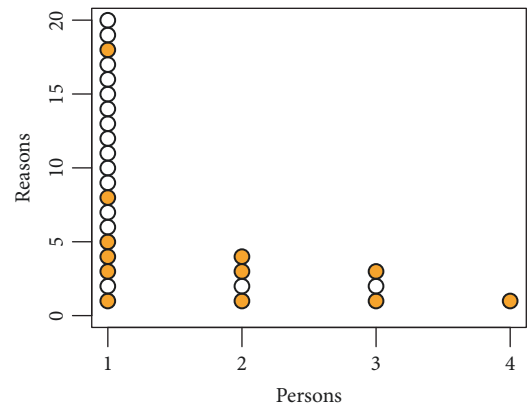


Figure 2: The reason space consists of reasons and the number of individuals that hold those reasons. We present a random individual who acknowledges reasons #1, #3, #4, #5, #8, and #18, which are held by 4, 3, 2, 1, 1, and 1 persons, respectively. The person–reasons covered by this person are shown in orange and represent this individual's representativeness, $\rho = 4 + 3 + 2 + 1 + 1 + 1 = 12$.

In other words, M does not change as it is sensitive to the structure of the person–reason network space.

3. Evaluation: Simulation Studies

3.1. Simulation 1. Consider a policy debate with two positions, for and against, with corresponding *pro* and *con* reasons. Individuals are aware of various reasons on both sides of the debate. The goal is to identify mediators in this space who are the best-recognizers-of-best-recognized reasons on both sides of the debate.

To simulate this, let there be $N = 100$ individuals in a policy debate on an issue (e.g., legalization of cannabis) where there is a universe of 10 possible distinct *pro* reasons and 10 possible distinct *con* reasons. Each individual samples a total of 10 reasons: 10β *pro* reasons and $10(1 - \beta)$ *con* reasons (rounded to the nearest integer). β then represents an individual's *bias*; a $\beta = 0.5$ represents an unbiased individual. In this simulation, an individual's β is uniformly sampled from $[0, 1]$.

We allow reasons to have a power law distributed probability p of being sampled, where $p \sim r^{-\gamma}$ with rank r and $\gamma = 2.5$. The precise value of γ is unimportant to the overall results except that for larger values of γ all individuals will produce the same reason and for small values of γ all reasons will be sampled uniformly. In such cases everyone is the best mediator (since everyone produces the same reasons) or the best mediator is the individual with the most number of reasons per position, since all reasons are equally represented. Thus, intermediate γ's are the most interesting formulation and also the formulation that best reflects the empirical data presented later.

Using individuals' samples, we produce a bipartite adjacency matrix where individuals are rows and reasons are columns. As described above, we then project this matrix onto individuals for the pro and con reason subspaces separately and then compute individuals' mediation centrality

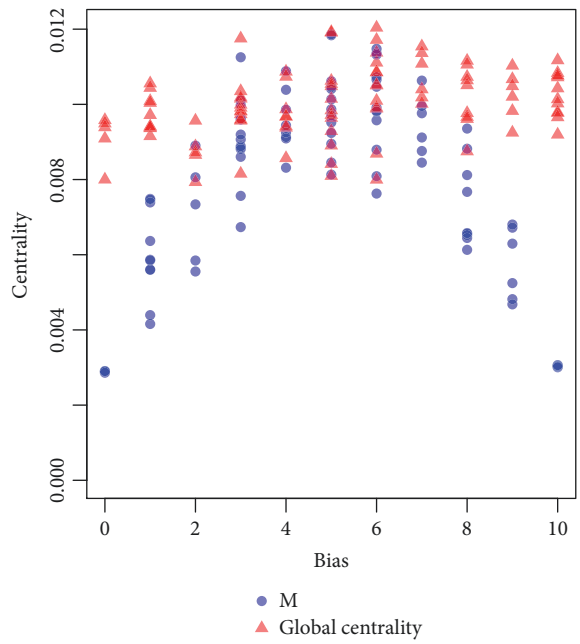


Figure 3: Individuals' centrality values and bias in Simulation 1. The centrality value of an individual is computed using two methods (mediation centrality and global centrality) and shown on the y-axis as a function of the individual's bias. Bias is shown here as the integer number of reasons on the pro side (i.e., 10β).

as the harmonic mean of their centrality values across both subspaces.

Figure 3 shows the relationship between global centrality (i.e., computed from the projection onto individuals based on the full bipartite network) and mediation centrality. Global centrality shows a limited ability to discriminate among individuals' different degrees of bias for or against the policy issue. Mediation centrality, on the other hand, captures the central intuition of mediation, whereby individuals with the least bias, $\beta = .5$, have the highest mediation centrality. Notably, however, the individual with the smallest bias is not necessarily the one with the highest mediation centrality. The variation in mediation for a given level of bias is a measure of the individual's ability to capture the most well-recognized reasons within a subgraph. Consider that individual's with no bias ($\beta = 0.5$) will produce an equal number of reasons from both sides of the issue, but these reasons may not be equally representative of the sides from which they are sampled. Therefore, more biased individuals can (up to a point) still be better recognizers of best-recognized reasons.

Figure 4 relates each simulated agent's representativeness ρ for pro and con spaces against each other and against the same agent's mediation centrality, M, and global centrality. The figure reveals a number of insights. Foremost, as to be expected, individuals who have higher representativeness in the con subspace have lower representativeness in the pro subspace. This is because the number of reasons produced is fixed and split across the subspaces. Secondly, individuals who are highly representative of one or the other of the two subspaces have lower mediation centrality (smaller sized dots; Figure 4(a)). Global centrality, on the other hand, is influenced by greater representativeness of either pro or con sides (Figure 4(b)). A comparison of mediation centrality with representativeness across both the pro and con spaces, $\widehat{\rho}$, shows that mediation centrality captures the intuition that mediators must acknowledge key ideas on each side of the issue (Figure 4(c)); global centrality lacks this property (Figure 4(d)). In addition, also note that the individual with the highest M is not the individual with the highest $\widehat{\rho}$.

Following the notion of a random walk through policy space, we can also verify that mediation centrality tracks the residence time of random walkers on each subgraph. Figure 5 shows the outcome of releasing 1,000 random walkers from each individual and tracking the residence times for each node in the network. Mediation centrality is again highest for the individuals who are least biased in their residence times (Figure 5(a)). Global centrality, in contrast, is less discriminating (Figure 5(b)). Comparing mediation centrality against the harmonic mean of residence times shows a close relationship between the two measures (Figure 5(c)). Global centrality, in contrast, does not show a clear relationship with the harmonic mean of residence times (Figure 5(d)).

3.2. Simulation 2. To further examine the characteristics of a good mediator, we ran a second simulation where there are 100 reasons in total, 80 against and 20 in favor of the policy issue; all other simulation details are identical to Simulation 1. Figures 6, 7, and 8 show the corresponding results. As expected, mediation centrality is relatively unaffected by the imbalance between the number of con and pro reasons and identifies individuals who are both unbiased and more representative; global centrality, in contrast, cannot capture the differences between the two subgraphs (Figure 6). Note also that an individual with the highest global centrality value has the lowest mediation representativeness and is strongly biased towards producing pro reasons (Figure 7(d)). Finally, mediation centrality again tracks the harmonic mean of random walker residence times, whereas global centrality fails to capture this (Figure 8).

The two simulations presented have highlighted the usefulness of mediation centrality in identifying good mediators across two sides of a policy issue. In the next section, we apply this measure to empirical data on actual policy issues.

4. Mediators in Adversarial Policy Networks: An Empirical Study

To test mediation centrality in a real-world context, we collected data for eight policy issues (Table 1). Participants in this study were asked to imagine that they would be moderating a discussion on a specific policy proposition and that in preparation for this meeting they should list all the possible reasons in favor or against the proposition they could think of that might come up in such a discussion. We call this task the "policy fluency task" following similar tasks in the category fluency literature, such as the animal and country fluency tasks, where individuals name all the animals or countries they can think of, respectively (e.g., [18, 27]).

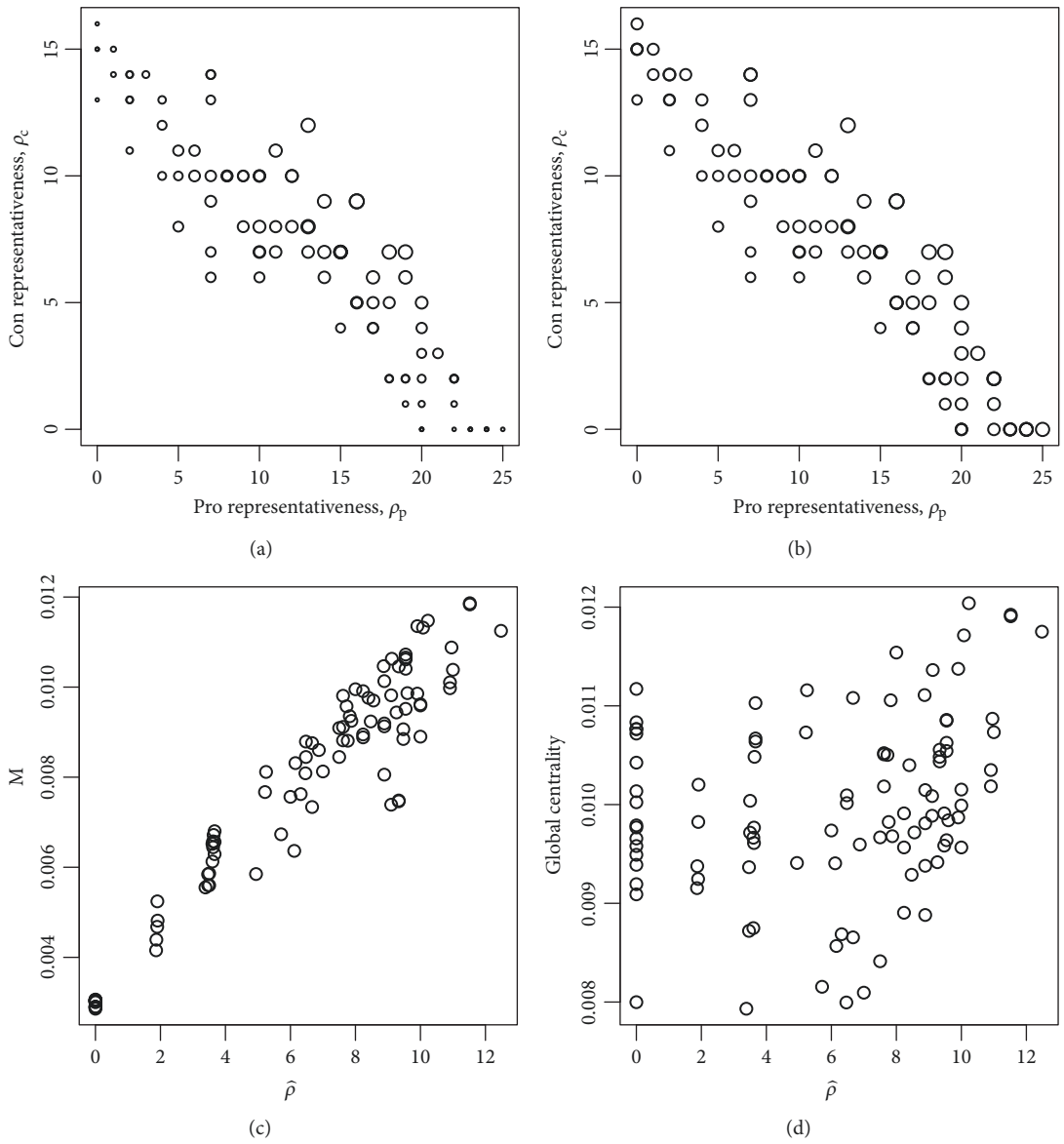


Figure 4: Comparison of mediation representativeness against mediation centrality and global centrality in Simulation 1. Each individual is represented by a dot. (a) Representativeness for pro and con reasons. Dot size represents mediation centrality. (b) Representativeness for pro and con reasons. Dot size represents global centrality. (c) Global representativeness against mediation centrality. (d) Global representativeness against global centrality.

Table 1: The 8 policy issues. Con and Pro: number of participants in favor of or against the policy; % Pro: percentage of participants in favor of the policy. The issues were framed in the context of the country in which the study was conducted (Switzerland).

Policy issue	Policy question: Should...	Con	Pro	% Pro
State-subsidized abortion	... the state subsidize abortions?	33	18	35
Bank bailouts	... the state bail out banks during an economic crisis?	20	26	57
Forced CO_2 reduction	... developing countries be forced to reduce CO_2 emissions?	29	20	41
Cannabis legalization	... the possession and consumption of cannabis be legalized?	29	20	41
Shortened naturalization	... the minimum years of residency for citizenship be reduced?	14	25	64
Surrogate legalization	... surrogate motherhood be legalized?	25	23	48
Public smoking ban	... public smoking be banned?	23	28	55
Euthanasia legalization	... medically assisted suicide by legalized?	22	31	58

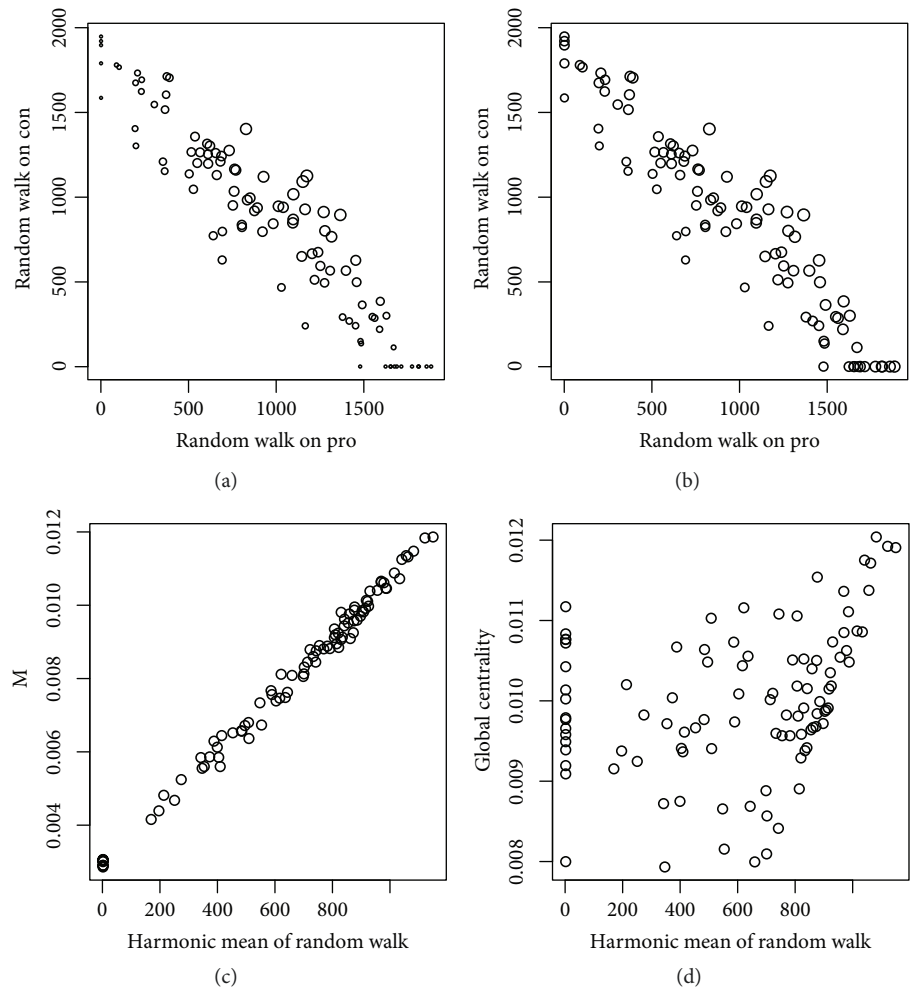


Figure 5: Comparison of mediation centrality and global centrality against random walk residence times in Simulation 1. Each individual is represented by a dot. (a) Random walk residence times for pro and con reasons. Dot size represents mediation centrality value. (b) Random walk residence times for pro and con reasons. Dot size represents global centrality. (c) Mediation centrality against random walk residence times. (d) Global centrality against random walk residence times.

4.1. Methods

4.1.1. Participants. Fifty-three participants (median age = 22; 40 females) were recruited at the University of Basel (Switzerland). As this experiment was a nonclinical study and did not involve any patients, according to Swiss federal law it did not require an in-depth evaluation and approval by a cantonal review board.

4.1.2. Materials and Procedure. We conducted a pilot survey to identify policy issues for which in our participant population there is nonnegligible support for both sides of an issue. Table 1 shows the eight issues used in the main study.

The primary data for this study are the reasons participants generated for each policy issue. Next to this primary data, we collected several additional variables that were not investigated in relation to mediation centrality. In the spirit of full disclosure we nevertheless report them below when describing the experiment. All instructions were in German; we present their English translations here. The experiment was programmed in E-Prime 2.

(1) To measure working memory capacity, participants completed an operation span task [28].

(2) Participants were asked to imagine that they would be moderating a discussion on a specific policy proposition (e.g., legalizing cannabis) and that their role was that of an impartial mediator. In preparation for this discussion they would list all the arguments (i.e., reasons) for and against the current proposition they could think of that other people might find important for deciding in favor of or against the policy proposition. For each of the eight policy issues (Table 1), participants were instructed to write down each reason they could think of using 3-4 words and submit it by pressing ENTER. Once they could not

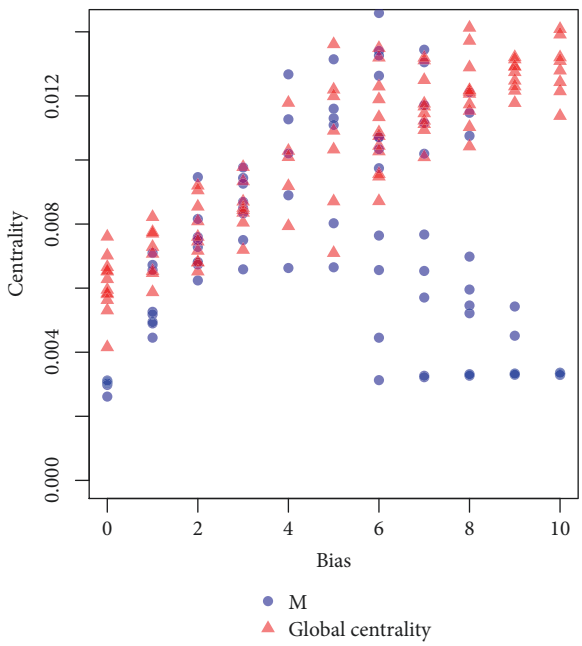


FIGURE 6: Individuals' centrality and bias values in Simulation 2. The centrality value of an individual is computed using two methods (mediation centrality and global centrality) and shown on the y-axis as a function of the individual's bias. Bias is shown here as the integer number of reasons on the pro side (i.e., 10β).

think of any more reasons, they proceeded to the next issue by pressing a button on the screen. Policy issues were presented in a new random order for each participant.

(3) For each of the eight policy issues participants indicated their own position (i.e., in favor or against the policy proposition).

(4) Participants indicated a set of demographic variables (age; gender; Swiss citizen status; smoking status; cannabis consumption).

(5) Several self-rating questions assessed participants' political stance. The first question asked participants to place themselves on the political left- versus right-wing spectrum by choosing a point on an analogue scale. Then for each of eight Swiss political parties participants indicated to which degree they agreed or disagreed with their political agenda. Participants choose a point on an analogue scale that ranged from "total disagreement" to "total agreement". The eight political parties were: Schweizer Volkspartei (SVP); Sozialdemokratishe Partei (SP); Freisinning demokratische Partei/FDP - Die Liberalen; Christlichdemokratische Volkspartei (CVP); Grüne Partei (GPS); Bürgerlich-Demokratische Partei (BDP); Grünliberale Partei (GLP); Evangelische Volkspartei (EVP).

(6) We assessed participants' self-reported ability for perspective taking based on the four items of the German version [29] of the Interpersonal Reactivity Index [30]. Participants indicated the degree to which four statements applied to them by choosing a point on an analogue bipolar scale that ranged from "does not apply at all" to "fully applies". In their original English formulation [30] the statements read: "I try to look at everybody's side of a disagreement before I make a decision." "I believe that there are two sides to every question and try to look at them both." "Before criticizing somebody, I try to imagine how I would feel if I were in their place." "When I'm upset at someone, I usually try to 'put myself in his shoes' for a while."

Three raters independently judged for each reason whether it was in support of (+1) or against (−1) the policy issue or whether they could not tell (0); we then summed the values and took the sign of the sum to indicate whether the reason was a pro or a con. Out of the total 1,778 reasons produced, 324 had no valence (i.e., a sum of zero). We excluded these as they often did not refer to coherent reasons.

For each issue, a fourth rater created overarching categories of reasons to which the produced reasons were then assigned; 7 of the remaining 1,454 reasons failed to be coded and were removed. The assigned categories were then used to compute the values in the adjacency matrices. For example, one individual wrote "murder of the fetus" and another wrote "it is murder" which were then classified under the same category "abortion is murder." An individual's representativeness was calculated by considering the number of unique reason categories for which a participant produced at least one reason. This was done to avoid inflated values of representativeness when a participant produced multiple reasons that all belonged to the same reason category.

After the two rating procedures, 1,447 reasons remained, which were used in all further analyses reported. The sum of the counts in Table 1 indicates how many out of the 53 participants produced at least one valid reason for the respective issue.

4.2. Results. Table 1 shows the number of individuals on each side of each issue. Our pilot survey aimed to identify policy issues for which there would be substantial support for either side in our participant population. Consistent with this goal, each of the issues showed nonnegligible support for both positions. These levels of polarization suggest that the issues used in the study represent a good test bed for investigating mediation centrality.

Figure 9 shows the bipartite adversarial reason networks for each issue. The networks each have one giant component, which shows that people on both sides of each issue tend to acknowledge reasons on both sides of the issue. Thus, even though these are controversial issues, participants were—at least partly—aware of the reasons the other side holds. This implies that identifying mediators as individuals who are the best-recognizers-of-best-recognized reasons is a plausible endeavor in this study.

Mediation centrality is useful to the extent that it varies across individuals in adversarial policy networks. Figure 10 shows that mediation centrality produces a clear ranking

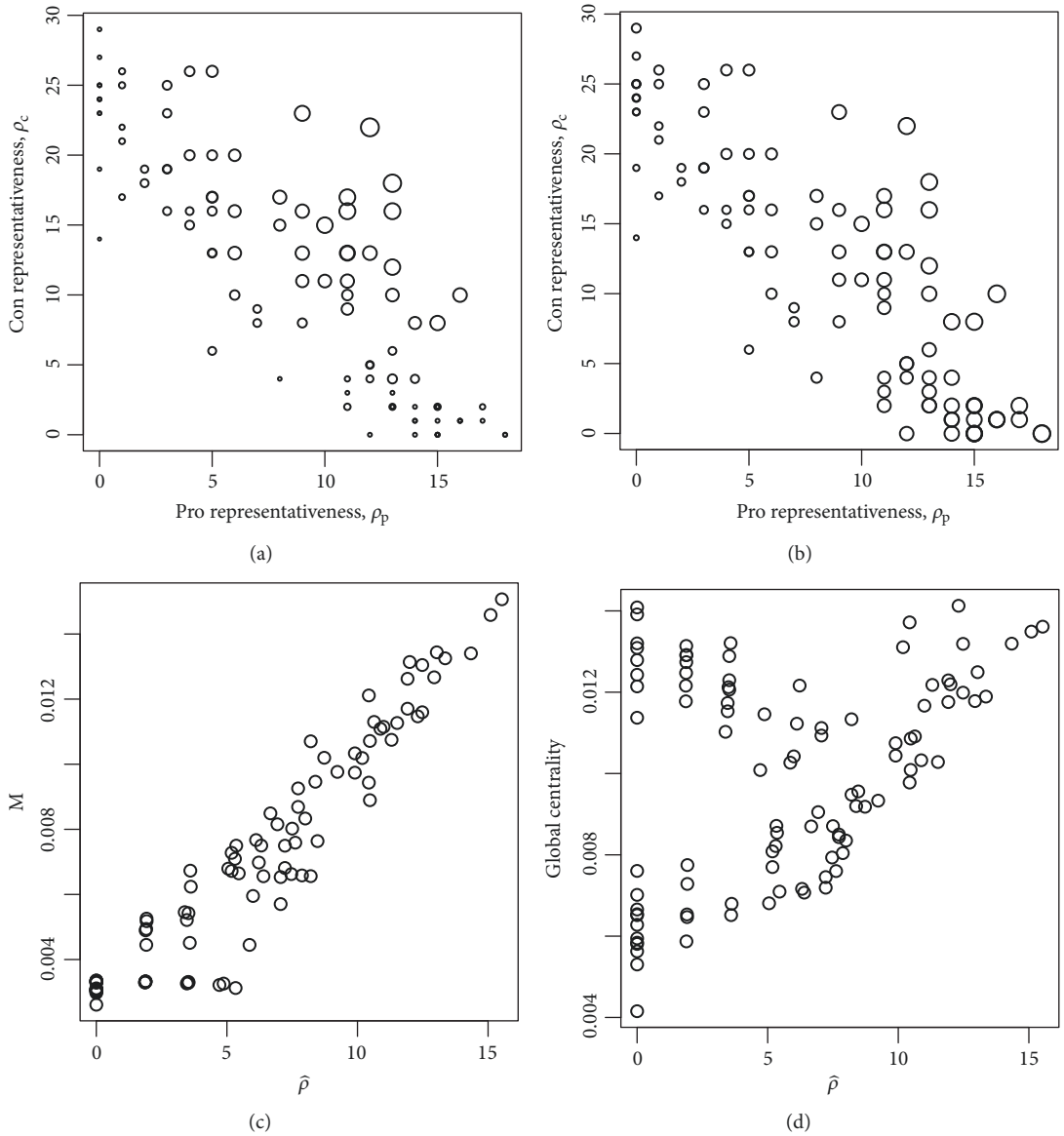


Figure 7: Comparison of mediation representativeness against mediation centrality and global centrality in Simulation 2. Each individual is represented by a dot. (a) Representativeness for pro and con reasons. Dot size represents mediation centrality. (b) Representativeness for pro and con reasons. Dot size represents global centrality. (c) Global representativeness against mediation centrality. (d) Global representativeness against global centrality.

of individuals within each of the eight very different policy issues. This is promising because this implies that even among controversial topics there is a range of adversarial understanding among people, or in other words, there are individuals who would be much better mediators than others.

Although the mere number of reasons produced by a participant is a rough proxy for the participant's mediation centrality, it is a poor direct substitute for mediation centrality. Figure 11 shows that although the highest mediation centrality corresponds in some cases to the individual with the most reasons produced, this is not always the case. Bailing out banks, shortened naturalization, public smoking ban, surrogate legalization, and forced CO_2 reduction demonstrate cases where producing the most reasons does not make one the best mediator.

Figure 12 compares mediation centrality M and mediation representativeness, $\widehat{\rho}$. The results are consistent with those shown in the simulations, indicating that individuals with higher $\widehat{\rho}$ also have higher M. However, they also show how in this real-world context $\widehat{\rho}$ can differ for individuals with the same M, such as the outliers in abortion and citizenship, which represent dyads separate from the respective giant components (networks not shown). Figure 13 shows similar results when comparing mediation centrality with the residence times of 1,000 random walkers starting at each individual in each subgraph. M is strongly correlated with

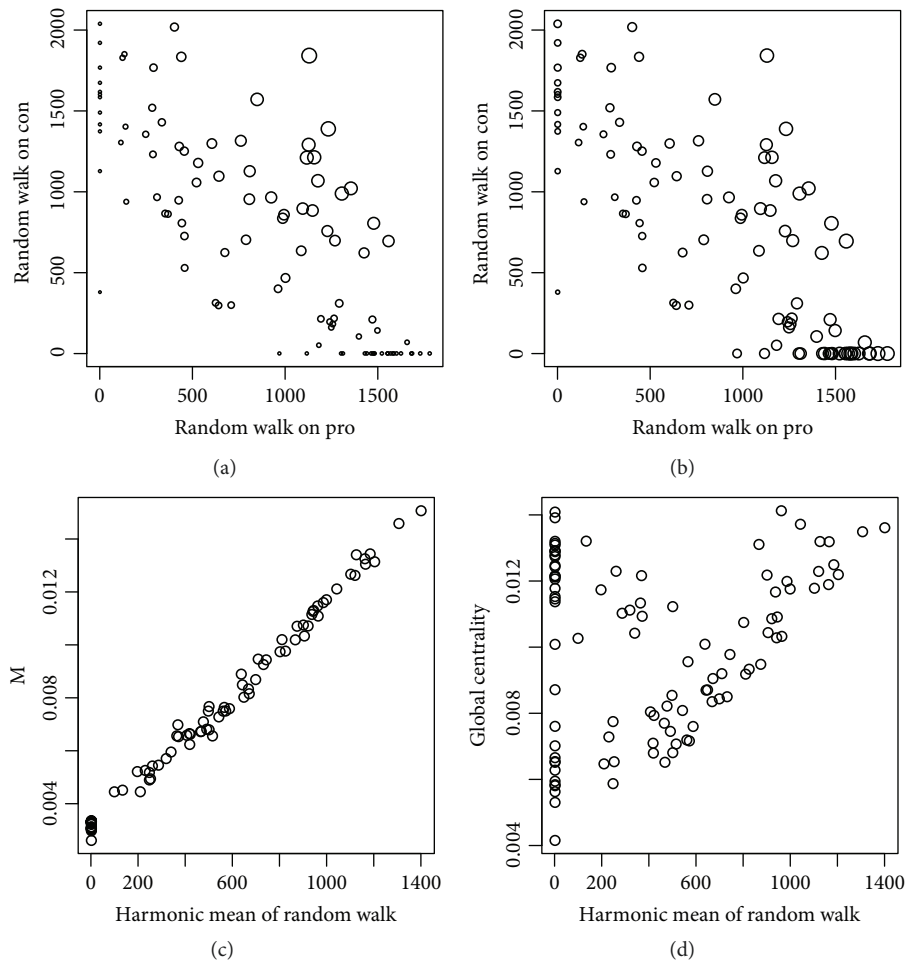


Figure 8: Comparison of mediation centrality and global centrality against random walk residence times in Simulation 2. Each individual is represented by a dot. (a) Random walk residence times for pro and con reasons. Dot size represents mediation centrality value. (b) Random walk residence times for pro and con reasons. Dot size represents global centrality. (c) Mediation centrality against random walk residence times. (d) Global centrality against random walk residence times.

the residence times of random walkers in the reason space. These results for mediation representativeness and random walk residence times demonstrate that mediation centrality conforms to our intuition of what a good mediator for these policy issues might look like: someone who is a best-recognizer-of-best-recognized reason.

5. General Discussion

The present article has two goals. The first and primary goal is to introduce a new measure for network scientists that capture an interesting and useful property of bipartite adversarial policy networks. As we show, mediation centrality has useful quantitative properties that can identify nodes in bipartite networks that may be particularly suited for certain tasks in adversarial settings. The second goal is to produce a measure of mediation that may be useful to social and cognitive scientists. Though this article focuses primarily on the former of these two goals, we nonetheless empirically demonstrated how mediation centrality is a meaningful measure for adversarial policy networks which should be useful in future studies that aim to provide more detailed quantification of the often rather qualitative conceptualization of mediation (see [10]).

In relation to complex networks, the mediation centrality measure we present, focusing on PageRank, is a particular instance of a more general family of possible measures of mediation centrality. Other centrality measures, such as closeness centrality or betweenness centrality, could be used in lieu of PageRank in the measure of mediation centrality we presented. However, we argue that for the task of mediation in adversarial policy networks, a measure of mediation that closely corresponds to social and cognitive processes is preferable to process-agnostic, descriptive measures, such as representativeness or other, more generic measures of network centrality. Nevertheless, such alternative measures are likely to be meaningful in other settings where they may closely correspond to other processes of interest. For example, when a bee colony tries to locate a position for their beehive that appropriately minimizes travel to resources

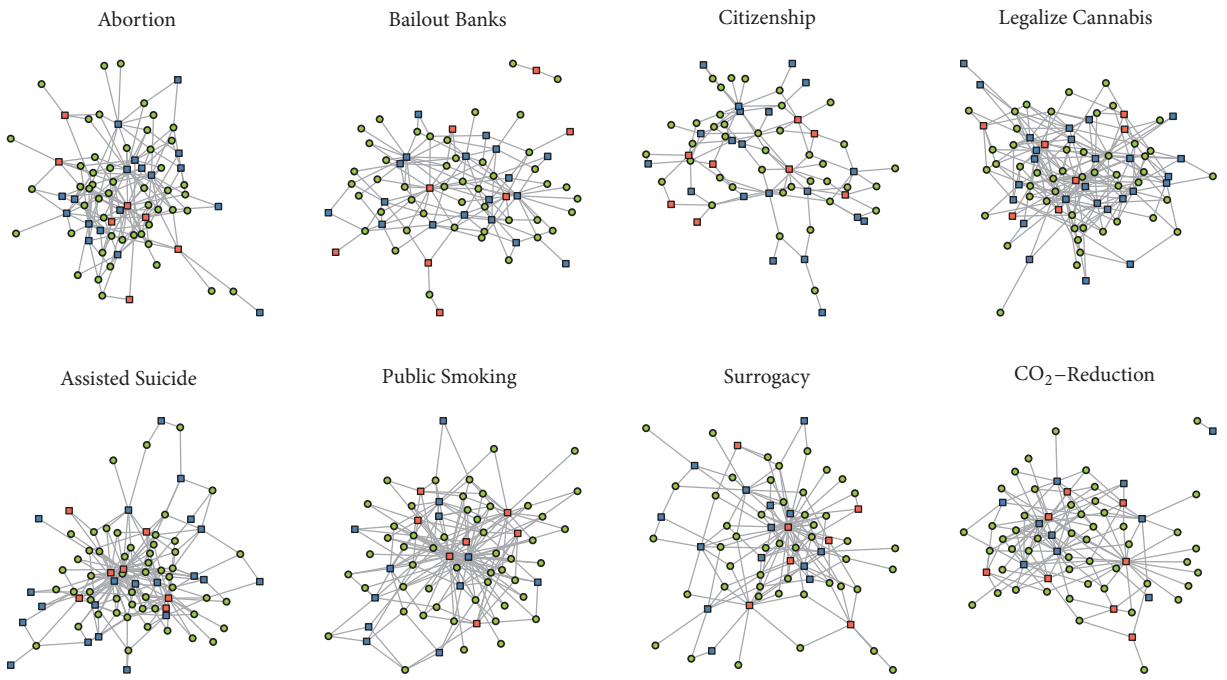


Figure 9: Bipartite adversarial policy networks for each issue, with reasons shown as boxes (red = con, blue = pro) and individuals shown as circles (green).

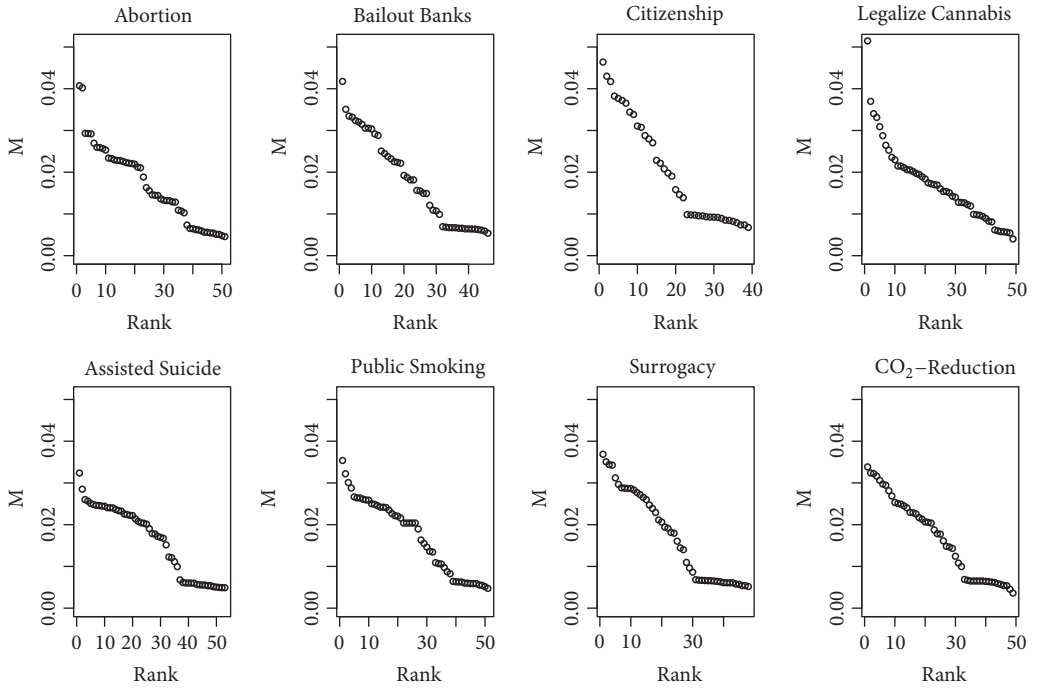


Figure 10: Participants' mediation centrality, ranked from largest to smallest, for each of the eight policy issues.

of different types, closeness mediation centrality may be highly appropriate. Mediation centrality can also be adapted to weighted subgraphs in relation to their relative importance by some other criteria, such as, for example, the number of people holding a particular position on the issue, or the value of different reasons (e.g., pollen over nectar in the beehive example above; [31]). Indeed, mediation centrality is a highly flexible approach for constructing quantitative measures and there is ample room for variations. For example, mediation centrality as proposed here is designed to measure mediation within a network dedicated to a given adversarial issue. But it may be valuable in the future to be able to quantify

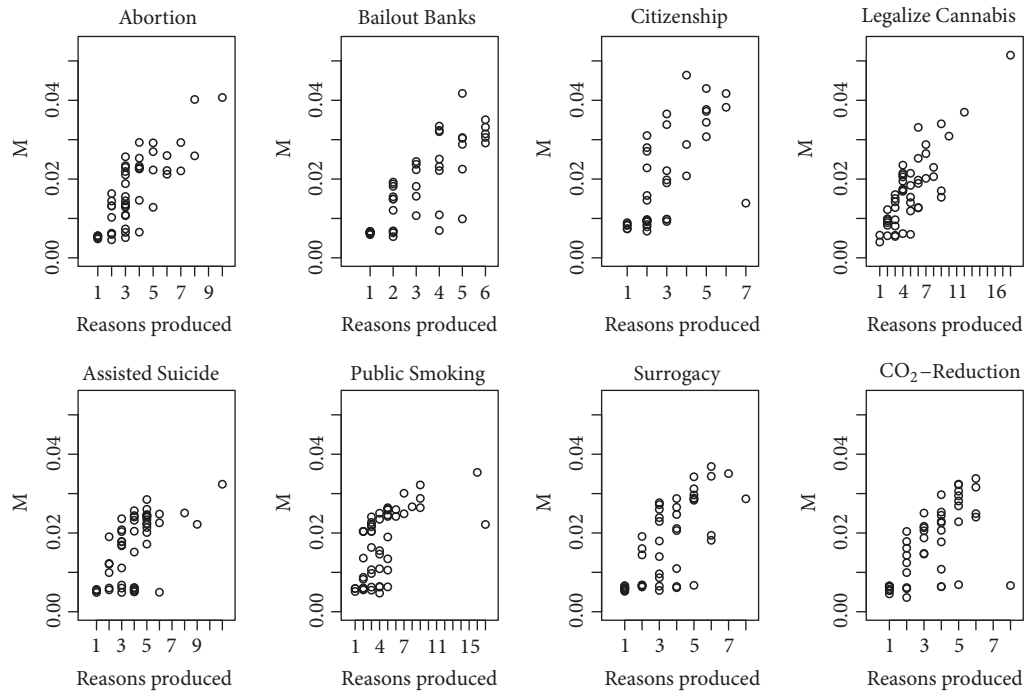


Figure 11: Mediation centrality M of an individual (y-axes) plotted as a function of the number of reasons produced by that individual, separately for each policy issue.

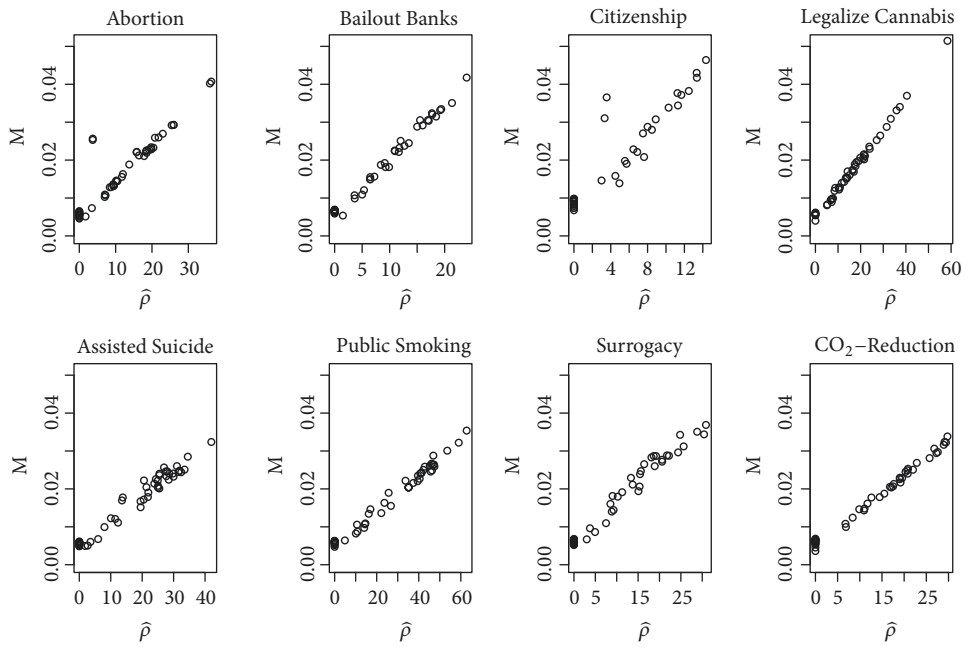


Figure 12: Mediation centrality M of an individual (y-axes) plotted as a function of mediation representativeness $\widehat{\rho}$ of that individual, separately for each policy issue.

the degree of mediation across multiple policy issues, with correspondingly different associative structures.

In addition, future investigations of mediation from a more social psychological perspective should focus on several factors that were not addressed here. Foremost, we did not capture the strength with which individuals held the positions they reported themselves as having. For example, we did not capture the strength with which an individual supported laws against public smoking, only that they did or did not. It would be useful to have a more graded measure of position, as one then could investigate whether individuals with high mediation centrality are also individuals who hold

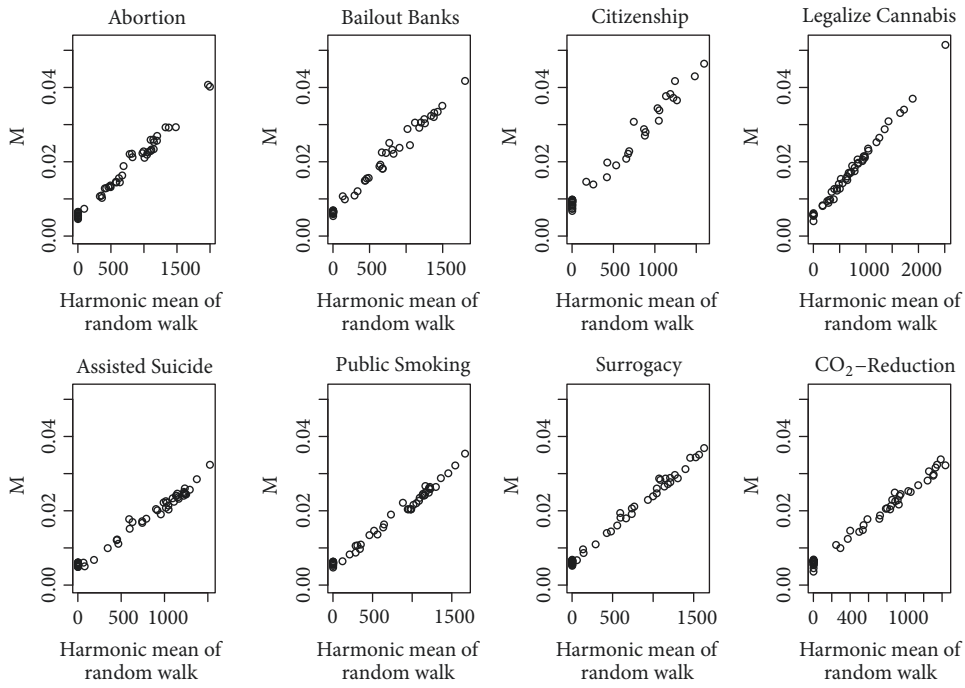


Figure 13: Mediation centrality M of an individual (y-axes) plotted as a function of the harmonic mean of random walk residence times (across both pro and con subgraphs), separately for each policy issue.

more moderate positions on these issues. This, of course, may not be the case. Good mediators by our measure may also—by virtue of their knowledge of what other individuals believe—be better persuaders. In this respect, future studies of mediation would also benefit by investigating the extent to which individuals with high mediation centrality can produce arguments that are more likely to lead to solutions recognized by both sides.

Future studies should also investigate the extent to which individuals believe the reasons they acknowledge to reflect legitimate arguments. Although we use a rather coarse measure of acknowledgement, merely involving the production of a reason, it may be that these reasons are acknowledged to different degrees. Both of the above issues could be adapted into future measures of mediation. Finally, it is important to note that mediation centrality, as we propose it here, is social-network agnostic. It solely focuses on *what you know* and not on *who you know*. Nonetheless, in many contexts, mediators may be most effective when they simultaneously know the relevant parties involved *and* recognize the best-recognized reasons held on alternate sides of the issue.

6. Conclusion

Individuals who can take the perspectives held by opposing sides of an issue and frame arguments in a way that both sides can agree on often help to generate better solutions during conflict resolution [8, 9, 32]. We apply this concept to mediators by extending perspective taking to a process-based account of mediation. This allows us to introduce mediation centrality, a metric for quantifying the mediation value of individuals in adversarial policy networks. Mediation centrality formalizes the notion of a good mediator in a collective cognitive representation of an adversarial policy space aggregated across multiple individuals and positions. Using simulations and empirical data from eight real-world policy issues, we show that mediation centrality follows the intuition of what it means to be a good mediator, and we further show how this outperforms other measures and captures the logic of a random walk over reason space.

Data Availability

The data and code used to support the findings of this study have been deposited in an Open Science Framework project (https://osf.io/nsd2r)

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

We thank Carmen Kaiser for coding the experiment, Theresa Schmitt for documenting the experiment, the CDS research assistants for recruiting participants, and Dania Esch, Eva Günther, Sebastian Lucht, and Sarah Turowski for coding participants' responses. This work was supported by the Royal Society Wolfson Research Merit Award (WM160074) and a Fellowship from the Alan Turing Institute (to Thomas T. Hills).

References

[1] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, "Tweeting from left to right: is online political communication more than an echo chamber?" *Psychological Science*, vol. 26, no. 10, pp. 1531–1542, 2015.

[2] D. Nikolov, D. F. M. Oliveira, A. Flammini, and F. Menczer, "Measuring online social bubbles," *PeerJ Computer Science*, vol. 1, p. e38, 2015.

[3] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, "Competition among memes in a world with limited attention," *Scientific Reports*, vol. 2, article 335, 2012.

[4] T. T. Hills, "The dark side of information proliferation," *Perspectives on Psychological Science*, 2018, in press.

[5] M. P. Fiorina and S. J. Abrams, "Political polarization in the American public," *Annual Review of Political Science*, vol. 11, pp. 563–588, 2008.

[6] F. Shi, Y. Shi, F. A. Dokshin, J. A. Evans, and M. W. Macy, "Millions of online book co-purchases reveal partisan differences in the consumption of science," *Nature Human Behaviour*, vol. 1, no. 4, article 0079, 2017.

[7] D. C. Mutz, "Cross-cutting social networks: testing democratic theory in practice," *American Political Science Review*, vol. 96, no. 1, pp. 111–126, 2002.

[8] A. D. Galinsky, W. W. Maddux, D. Gilin, and J. B. White, "Why it pays to get inside the head of your opponent: the differential effects of perspective taking and empathy in negotiations," *Psychological Science*, vol. 19, no. 4, pp. 378–384, 2008.

[9] L. E. Drake and W. A. Donohue, "Communicative framing theory in conflict resolution," *Communication Research*, vol. 23, no. 3, pp. 297–322, 1996.

[10] M. Deutsch, P. T. Coleman, and E. C. Marcus, *The Handbook of Conflict Resolution: Theory and Practice*, Jossey-Bass, San Francisco, CA, USA, 2nd edition, 2006.

[11] G. Ku, C. S. Wang, and A. D. Galinsky, "The promise and perversity of perspective-taking in organizations," *Research in Organizational Behavior*, vol. 35, pp. 79–102, 2015.

[12] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Networks*, vol. 29, no. 4, pp. 555–564, 2007.

[13] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.

[14] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[15] M. Newman, *Networks*, Oxford University Press, Oxford, UK, 2 edition, 2018.

[16] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[17] T. Hobbes, *Leviathan*, Oxford University Press, Oxford, UK, 1998, Original work published 1651.

[18] T. T. Hills, M. N. Jones, and P. M. Todd, "Optimal foraging in semantic memory," *Psychological Review*, vol. 119, no. 2, pp. 431–440, 2012.

[19] T. L. Griffiths, M. Steyvers, and A. Firl, "Google and the mind: predicting fluency with PageRank," *Psychological Science*, vol. 18, no. 12, pp. 1069–1076, 2007.

[20] J. R. Norris, *Markov Chains*, No. 2, Cambridge University Press, Cambridge, UK, 1998.

[21] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1, pp. 107–117, 1998.

[22] J. L. Austerweil, J. T. Abbott, and T. L. Griffiths, "Human memory search as a random walk in a semantic network," in *Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012, NIPS 2012*, pp. 3041–3049, USA, December 2012.

[23] J. Borge-Holthoefer and A. Arenas, "Semantic networks: structure and dynamics," *Entropy*, vol. 12, no. 5, pp. 1264–1302, 2010.

[24] Y. Ding, E. Yan, A. Frazho, and J. Caverlee, "PageRank for ranking authors in co-citation networks," *Journal of the Association for Information Science and Technology*, vol. 60, no. 11, pp. 2229–2243, 2009.

[25] E. Estrada and J. A. Rodriguez-Velazquez, "Subgraph centrality in complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 5, Article ID 056103, 2005.

[26] T. Russell and T. Reimer, "Using semantic networks to define the quality of arguments," *Communication Theory*, vol. 28, no. 1, pp. 46–68, 2018.

[27] T. T. Hills and E. Segev, "The news is American but our memories are... Chinese?" *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1810–1819, 2014.

[28] N. Unsworth, R. P. Heitz, J. C. Schrock, and R. W. Engle, "An automated version of the operation span task," *Behavior Research Methods*, vol. 37, no. 3, pp. 498–505, 2005.

[29] C. Paulus, "Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index," *The Saarbrueck Personality Questionnaire on Empathy: Psychometric Evaluation of the German Version of the Interpersonal Reactivity Index*, pp. 1–11, 2009, http://hdl.handle.net/20.500.11780/3343.

[30] M. H. Davis, "Measuring individual differences in empathy: evidence for a multidimensional approach," *Journal of Personality and Social Psychology*, vol. 44, no. 1, pp. 113–126, 1983.

[31] T. D. Seeley, *The Wisdom of The Hive: the Social Physiology of Honey Bee Colonies*, Harvard University Press, 2009.

[32] K. E. Kemp and W. P. Smith, "Information exchange, toughness, and integrative bargaining: The roles of explicit cues and perspective-taking," *International Journal of Conflict Management*, vol. 5, no. 1, pp. 5–21, 1994.

Hindawi
Complexity
Volume 2019, Article ID 6979830, 21 pages
https://doi.org/10.1155/2019/6979830

Research Article

Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation

Karl D. Neergaard [1,2] **and Chu-Ren Huang** [2]

[1] *Laboratoire Parole et Langage, Aix/Marseille University, Aix-en-Provence 13604, France*
[2] *Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong*

Correspondence should be addressed to Karl D. Neergaard; karlneergaard@gmail.com

Received 29 June 2018; Revised 15 October 2018; Accepted 5 November 2018; Published 23 April 2019

Guest Editor: Cynthia Siew

The purpose of this study was to construct, measure, and identify a schematic representation of phonological processing in the tonal language Mandarin Chinese through the combination of network science and psycholinguistic tasks. Two phonological association tasks were performed with native Mandarin speakers to identify an optimal phonological annotation system. The first task served to compare two existing syllable inventories and to construct a novel system where either performed poorly. The second task validated the novel syllable inventory. In both tasks, participants were found to manipulate lexical items at each possible syllable location, but preferring to maintain whole syllables while manipulating lexical tone in their search through the mental lexicon. The optimal syllable inventory was then used as the basis of a Mandarin phonological network. Phonological edit distance was used to construct sixteen versions of the same network, which we titled phonological segmentation neighborhoods (PSNs). The sixteen PSNs were representative of every proposal to date of syllable segmentation. Syllable segmentation and whether or not lexical tone was treated as a unit both affected the PSNs' topologies. Finally, reaction times from the second task were analyzed through a model selection procedure with the goal of identifying which of the sixteen PSNs best accounted for the mental target during the task. The identification of the tonal complex-vowel segmented PSN (C_V_C_T) was indicative of the stimuli characteristics and the choices participants made while searching through the mental lexicon. The analysis revealed that participants were inhibited by greater clustering coefficient (interconnectedness of words according to phonological similarity) and facilitated by lexical frequency. This study illustrates how network science methods add to those of psycholinguistics to give insight into language processing that was not previously attainable.

1. Introduction

The meeting of network science and the study of phonological processing has allowed for the examination of the mental lexicon according to mathematical principles that have both theoretical and methodological import. Researchers have used what are known as phonological networks, in which words (nodes) are connected to other words (edges) based on phonological similarity. Phonological networks have been used in basic research to examine speech processing during word recognition [1, 2], word production [3], word learning [4], and working memory [5]. Most recently, they have been applied to the study of speech pathologies in the examination of both aphasic speech [6] and stuttering [7, 8].

Common among the phonological networks that have been examined thus far is that phonological similarity is measured at the level of the phoneme. This is due to the relational parameter between nodes being defined by phonological edit distance, wherein two words are "neighbors" if they differ through the addition, deletion, or substitution of a single phoneme [9]. A given node's degree is thus the total number of words that are immediate neighbors, most commonly referred to as phonological neighborhood density [10]. One possible problem with using the phoneme as the basic unit between words is its generalizability to non-European languages. While English has gained attention in modeling network topologies of phonology [9–13], little has been done outside English. The two studies to date [14, 15]

implemented the single-phoneme edit metric. Yet, is a one size fits all approach appropriate cross-linguistically? This is an especially pertinent question in light of Mandarin Chinese. Mandarin not only has unique lexical features that set it apart from the languages studied to date but has long enjoyed a debate as to both its phonological annotation and syllable segmentation, i.e., two aspects that would likely affect network dynamics.

Mandarin has become a recent focus in the psycholinguistic literature due to a set of linguistic features that test the limits of models of speech processing previously developed for European languages. Perhaps the most unique is the status of the syllable, which is tonal, of equivalent size to the primary orthographic unit, and highly homophonic. Unlike English or Dutch, which both have over 10,000+ syllables, the Mandarin syllable inventory is small, featuring ~1,300 syllables plus tone and ~400 without tone. Excluding a select number of high-frequency lexical items that do not regularly carry tone, each syllable carries one of four tones: tone 1 (high level pitch, 55), tone 2 (low rising pitch, 35), tone 3 (low dipping pitch, 214), or tone 4 (high falling pitch, 51). Aside from the dialectal phenomena known as erhua [16] in which the character 儿 (*er2*) is added to another character yet pronounced as a single syllable (玩儿, *wan2 er2* = *war2*), each syllable in the inventory matches one or more Chinese characters. Mandarin has been shown to be largely disyllabic in nature; in fact it has been calculated that two-thirds of all Mandarin words consist of two characters [17, 18]. Yet, characters that do not exist as monosyllabic words, meaning they only exist in multisyllabic words, are still lexical items that contribute to the count of homophone neighbors. In context, the same roughly 1,300 tonal syllables service all lexical combinations from monosyllabic to multisyllabic words. This leads to a homophone density (i.e., the number of homophone neighbors a given word has) of up to 48 when tone is considered [19]. To put this in context, 11.6% of Mandarin words have homonyms, compared to 3.15% in English [20]. High homophony has been shown to lead to lexical competition in spoken word recognition, as seen by slower reaction times, and lower accuracy [19, 21]. This is uniquely important to Mandarin given the relation of orthography to the syllable.

Researchers have used two methods to describe how segmental units comprise a syllable in Mandarin. One method recognizes a maximum of 4 segments, CGVX, such that C represents initial consonants, G medial glides, the V monophthongs, and the final X the second part of a diphthong, or a final consonant. Early accounts proposed segmentation schemas based on the constituents of the rime or whether the medial glide constituted a unique phonological role within the syllable: C_GVX [22]; C_G_VX [23, 24]; CG_VX [25]; and CG_V_X [26, 27]. Note that here an underscore denotes a separation between phonological units. The second method of describing the Mandarin syllable collapses all glide and vowel information, leaving a maximum of 3 units: C_V_C [28]. The methods used to arrive at the various schemas, and whose evidence has informed the creation of syllable inventories, come through production tasks that have participants read sentences so as to measure syllable durations [29–31], or produce phonological neighbors in rhyming games [25, 32, 33]. More recent approaches depart from these methods to instead investigate segmentation as a product of either perception or production.

O'Séaghdha and colleagues [34, 35] hypothesized that the first phonological units available for selection below the level of the word or morpheme, titled proximate units, correspond to nontonal syllables in Mandarin. Their thesis was that unit sizes would vary across languages, granting phonemes and clusters of phonemes in Indo-European languages such as English, while larger units such as morae in Japanese. Speech error analyses have supported this trend, such that in English the dominant unit size is segmental [36, 37] and in Mandarin syllabic [38–40]. For speakers from alphabetic languages, like English and Dutch, sensitivity to syllable onsets between two lexical items has been documented in numerous studies and across multiple paradigms [35, 41–46]. These studies show that prior preparation to segmental units shared between lexical primes and target lexical items speeds production of the target word, implying that temporary storage occurs for segmental information. A corresponding series of priming studies have shown syllabic priming results yet no significant onset priming with Chinese orthography in the implicit priming [35, 47] and masked priming tasks [48–50]. To counter the syllable bias of Chinese characters, similar studies were conducted with picture [49, 51] and auditory stimuli [51]. Supporting evidence for the proximate unit has also been advanced in priming studies with both speakers of Cantonese [52–54] and Japanese [55–58]. While no syllable schema was explicitly proposed by the authors related to the proximate unit proposal, their statement that the primary unit to be selected is nontonal suggests that either a nontonal unsegmented schema is the target (CGVX), or its tonal counterpart (CGVX_T) seeing as the syllable is combined with tone prior to production.

To stand in contrast to speech production studies is a growing body of evidence to support the claim that Mandarin speakers, during speech recognition, process segmental information incrementally and in parallel with tonal information. Differential processing between lexical units was analyzed within a picture-word matching paradigm with both ERP [59–61] and eye-tracking [62]. Malins and colleagues found that whole-syllable mismatches did not produce effects greater than those found with individual components, mirroring results previously found in English [63]. This has motivated their claim that processing was segmental, an assertion not entirely supported in [61], which found greater evidence for syllable-level processing. One important difference in the latter study however is the fact that they also used Chinese characters during the presentation of their picture stimuli. Their results have likely an effect due to the activation of syllable-sized orthography. One limit to the claims put forward by Malins and colleagues, which implies words reside within a tonal fully segmented schema (C_G_V_X_T), is that they did not feature mismatch pairs according to glides, or the X unit [59, 60, 62]. Thus, to date these studies provide evidence for the tonal complex-onset/rime schema (CG_VX_T).

The current study began from the ground up through the creation of a novel phonological annotation system, also concurrently referred to as a syllable inventory. The creation of a novel inventory was necessary because (1) differences between existing inventories [23, 64–66] can be quite substantial, and (2) none of the existing inventories were made specifically with phonological similarity in mind. The novel inventory was constructed through participant-elicited phonological neighbors in two phonological association tasks. Phonological association tasks have been used with both nonword [67–71] and word stimuli [2, 72, 73] and provide information pertinent to syllable segmentation and the units being manipulated in that participants are asked to create minimal pairs. Minimal pairs have long played an important role in the identification of phonological units [74, 75] because of their ability to distinguish allophones from phonemes. In both tasks, we identified the respective salience of syllabic and tonal units according to edit distance (the difference between two words in number of segmental units), edit type (whether the manipulation between one word and the next is made through the addition, deletion, or substitution of a segmental unit), and edit location (i.e., the structural unit that a manipulated segment corresponds to in a fully segmented syllable: C_G_V_X_T).

In Experiment 1 we used our participants' productions to evaluate 2 existing annotation systems. We then constructed a new annotation system based on the gaps where either inventory disagreed with our participants' productions. In Experiment 2 we then validated which system optimally represented Mandarin phonology in light of phonological similarity.

The optimal annotation system was then used to construct sixteen phonological networks (8 with tone and 8 without tone), each built from an existing proposal, or suggested permutation, of Mandarin syllable segmentation. To avoid confusion between terms such as network, or schema, we introduce the term phonological segmentation neighborhood (PSN). Each of the sixteen PSNs is a representation of the same lexicon built upon a different schematic representation. While they all share the same lexical items, they differ in what constitutes neighbors. For example, given the phonological word *xiang4*, (as written in Mandarin Romanization, a.k.a. pinyin) the neighbors for the tonal fully segmented PSN (C_G_V_X_T) differ from its nontonal equivalent (C_G_V_X) by three items (*xiang1*, *xiang2*, and *xiang3*). Differences in degree and other topological network statistics accordingly arise between each PSN due to combinations of segmental units and whether or not tone is included in the calculation of similarity between words. The topological network characteristics of each PSN were analyzed similar to the analysis of [14].

Finally, an analysis was performed to identify which of the possible schematic representations of the Mandarin phonological mental lexicon best represented the task demands. We proceeded under the modeling assumption that a given target word within the metrical frame of the Levelt model [76], or the phonological representation frame of the Dell model [77], would share the same segmentation properties as the words they are connected to in long-term memory. We exploited the differences in local network features (i.e., word level) between the sixteen PSNs in a model selection procedure. Reaction times from the second association task were fitted to multiple lexical statistics per PSN with the goal of identifying which best represented the underlying mental procedure of retrieving phonological neighbors. Due to previous findings in Mandarin speech production studies, our hypothesis was that an unsegmented syllable, either tonal (CGVX_T) or nontonal (CGVX), would be identified in our modeling procedure and that like [70], greater density, as calculated on the network's degree, would facilitate mental search.

2. Experiment 1

2.1. Methods

2.1.1. Participants. Thirty-four native-Mandarin speaking participants (Female: 21; Age: M, 24.74; SD, 5.29) took part in this experiment. None of the participants reported a history of speech or hearing disorders. Prior to the experiment, participants were asked to complete a short biographical survey. Contents of the survey included, besides age and sex, the name of their home province, self-rated spoken fluency on a scale of 1 (beginner) to 10 (native speaker) in English (M: 6.26; SD: 1.11), and other Chinese languages/dialects and/or other non-Chinese languages. From their home province, we classified the speakers into two groups based on whether the region was traditionally a Mandarin (Guanhua) speaking region (Guanhua: 21; non-Guanhua: 13). To represent increased competition between similar Chinese languages/dialects, we summed the number of Chinese languages/dialects for all self-rated values from 3-10. This gave us a value that roughly reflects the number of Chinese languages/dialects (M: 2.14; SD: 0.56) that would have words similar to our target Mandarin stimuli. All participants reported native-level proficiency in Mandarin.

The Hong Kong Polytechnic University's Human Subjects Ethics Subcommittee (reference number: HSEARS20140908002) reviewed and approved the details pertinent to all experiments conducted in this study prior to beginning recruitment. The participants gave their informed consent and were compensated with 50HKD for their participation.

2.1.2. Stimuli. The material consisted of 155 Mandarin monosyllabic words, which can be seen in Table 1. The stimuli belonged to three groups according to the phase in which they were given to the participants: Example minimal pairs, 32; Practice, 10; Test, 113. A female speaker from the Beijing area produced all of the stimuli by speaking target monosyllables at a normal speaking rate 5 times into a high-quality microphone. Clearly produced items that were closest to the group mean in length were chosen. The pronunciation of each monosyllabic word was verified through transcriptions done by native-Mandarin speaking volunteers. Stimuli that did not have full agreement between transcribers were rerecorded and rerated until all stimuli were verified by at least 10 volunteers. All stimuli were edited using Audacity 2.1.2 and were 415ms in length.

Table 1: Experiment 1 stimuli.

Example pairs							
bi3~bian3		*chi3~zi3*		*diu1~di1*		*fo2~fei2*	
huang2~hua2		*ka3~kua3*		*lie4~luo4*		*mian2~miao2*	
miu4~you4		*nie4~nue4*		*ou1~sou1*		*piao4~pao4*	
ran2~rang2		*shan1~shan4*		*tian2~tuan2*		*zhe4~zhen4*	
Practice							
	bing1	*cai2*	*chu1*	*fa3*	*guai4*		
	mei2	*reng2*	*shuo1*	*song4*	*zui4*		
Test							
Base Rime	+C	+C		Base Rime	+C	+C	
a1	*ba3*	*ma1*		*wang4*	*kuang2*	*zhuang1*	
ai4	*gai1*	*zai4*		*wei2*	*chui1*	*tui3*	
an4	*nan2*	*san1*		*wen4*	*hun4*	*zun1*	
ang2	*shang4*	*tang3*		*weng1*			
ao4	*kao4*	*lao3*		*wo3*	*cuo4*	*huo2*	
e4	*che1*	*de2*		*wu3*	*fu4*	*ru2*	
ei4	*hei1*	*pei2*		*ya4*	*dia3*	*xia4*	
en1	*fen4*	*gen1*		*yan3*	*tian1*	*qian2*	*pian4*
	deng3	*zheng4*		*yang3*	*niang2*	*xiang3*	
	ren2	*sen1*		*yao4*	*tiao2*	*xiao3*	
er2				*ye2*	*bie2*	*jie1*	
	di4	*li3*	*ni3*	*yi1*	*ji1*	*qi3*	
	lin2	*pin1*		*yin1*	*jin4*	*xin1*	
	ming2	*ting1*		*ying2*	*jing3*	*qing3*	
	ri4	*si3*		*yong4*	*qiong2*	*xiong2*	
	bo1	*mo2*		*you3*	*niu2*	*qiu2*	
	hong2	*cong2*		*yu3*	*lv4*	*nv3*	
ou4	*hou4*	*rou4*		*yuan2*	*juan4*	*quan2*	
wa1	*gua4*	*zhua1*		*yue4*	*jue2*	*xue2*	
wai4	*kuai4*	*shuai4*		*yun4*	*lun2*	*xun2*	
wan2	*guan3*	*suan4*					

The example minimal pairs were exposed to the participants prior to the practice phase of the experiment. The idea of providing auditory examples of sound similarity came about during piloting. Upon given instructions to create minimal pairs or similar sounding syllables, our pilot participants were by and large unsure of what constituted similarity. Luce and Large [71] avoided this possible pitfall by providing their participants the one-phoneme difference rule, while Wiener and Turnbull [70] made it explicit which segment was to be manipulated in three of their four experimental blocks. We chose to provide example pairs because we did not want to bias our participants towards a tonal fully segmented syllable (C_G_V_X_T); however, it was not possible to provide a perfectly even example per each segmentation schema specifically because syllables can be interpreted in multiple ways depending on the number of units in the syllable, or the interpretation of the segments within the syllable. For example, the syllable pairs *bi3~bian3* can be interpreted as both C_GVX_T and CG_VX_T with the Z&L inventory (/pi^{214}/, /pian^{214}/), while only representing C_GVX_T with the Lin inventory (Lin: /pi^{214}/, /pjɛn^{214}/). Of the 17 example pairs presented to our participants, 7 consisted of a single-segment manipulation according to both inventories (*chi3~zi3*; *ou1~sou1*; *miu4~you4*; *nie4~nue4*; *ka3~kua3*; *piao4~pao4*; *shan1~shan4*), while 9 consisted of multiple segment manipulations according to either Lin or Z&L (*fo2~fei2*; *ran2~rang2*; *zhe4~zhen4*; *huang2~hua2*; *tian2~tuan2*; *lie4~luo4*; *mian2~miao2*; *bi3~bian3*; *diu1~di1*).

The test stimuli were created with the goal of representing all syllable structures in the Mandarin language. This was done by adding two lexical items per each base rime syllable from the syllable inventory through the addition of a consonant, regardless of lexical tone. For example, the addition of the consonants, /k/ and /ts/, to the nontonal base syllable /ai/ produced *gai1*, *zai4*, and *ai4*, respectively. Noteworthy about the choice of stimuli, which can be seen in Table 1, are some peculiarities due to the nature of the syllable inventory. First, the base rime syllables, *er2*, and *weng1*, do not have corresponding initial consonant phonological neighbors. Conversely, the pinyin syllables ending in "*eng*" (Lin and Z&L: /əŋ/), such as *deng3*, and *zheng4*; "*i*" as found in *ri4, si3* (Lin: /ɻ̩/ and Z&L: /ɿ/); and "*ong*" (Lin and Z&L:

/uŋ/), located in *hong2* and *cong2*, do not have corresponding base rime syllables. Next, based on phonotactic concerns and a high incidence of certain onsets, we included extra entries. The pinyin onset consonants "*j*" /tɕ/, "*q*" /tɕʰ/, and "*x*" /ɕ/ only cooccur with a small range of rimes, most notably accompanied by the three glides (Lin: /j, ɥ, w/; Z&L: /i, y, u/). Their greater occurrence with glide rimes meant that choosing them was unavoidable and would consequently lead to their overrepresentation in the stimuli set ("*j*" /tɕ/: 6 stimuli; "*q*" /tɕʰ/: 6 stimuli; "*x*" /ɕ/: 7 stimuli). We thus added onsets with lower occurrence for the pinyin rimes ending in "*i*" (*di4, li3, ni3*), "*in*" (*lin2, pin1*), "*ing*" (*ming2, ting1*), and "*en*" (*ren2*, *sen1*).

2.1.3. Procedure. Seated in a quiet room in front of a computer running E-Prime 2.0 [78] and wearing headphones equipped with an adjustable microphone, each participant was exposed to three phases: pretraining, practice, and test. Prior to beginning the experiment participants were instructed to not produce nonitems, which included syllables that do not correspond to an existing Chinese character. For the pretraining phase, participants were told to listen and not respond as they were exposed to 17 word pairs as examples of similar sounding syllables. Each pair was presented according to the same procedure: an auditory stimulus was presented during a blank screen that lasted the word's duration followed by a slide that read "听起来像" (sounds like) for 500ms, that was then immediately followed by its minimal pair during a blank screen that lasted the duration of the stimulus. Between each pair, a dark grey slide that featured, "…", in the center of the screen remained for 2000ms.

For the practice phase participants were told to produce a similar sounding syllable for each of the 10 items. Each stimulus was presented on a blank screen with no time limit. Participants were told that their spoken responses would advance the next trial by activating the PST Serial Response Box. A pause of 500ms followed each participant's response, followed by a slide that read "下一个词" (next word) for 500ms before the next trial. The test phase followed the same procedure for all 113 randomized test items. The entire task took an average of 15 minutes to complete. The audio was recorded on a second computer using Audacity 2.0.6 for offline analysis.

2.2. Results and Discussion. Two native-Mandarin speaking volunteers transcribed into pinyin the participants' spoken productions, with an agreement rate of 93%. A third transcriber resolved disagreements or classified unresolved items as nonitems. The pinyin responses were then translated into a sampa (ascii phonological transcription) that accommodated both the Lin and Z&L syllable inventories.

Our participants responded with large numbers of legal syllables that corresponded to existing Chinese characters, but were not monosyllabic words. We did not discount these items due to their qualification as lexical items. The online dictionary www.zdic.net [79] was used to classify nonitem status by identifying whether a given syllable corresponded to an existing Chinese character. Zdic.net, which includes definitions and pronunciations for 75,983 characters, has been used as a resource in the disambiguation of out-of-vocabulary words in several studies [80–83].

Missing (67), identical responses (138), and nonitems (260) were excluded, accounting for 12.16% of the total 3,842 observations.

2.2.1. Syllable Inventory Creation. In the current section, we detail the creation of a novel syllable inventory through the use of our participants' productions. It is first important to note that neither the Lin nor Z&L inventory, which will be used in the process described below, was constructed specifically according to phonological similarity. While Lin's inventory was informed through phonetic analysis, the Z&L inventory was created to be used in computational models of lexical processing. They are valuable for the current purpose because they have critical differences. The two inventories differ according to glides, as can be seen in syllables such as *ying2*, and *qing3* (Lin: /jəŋ^{35}/, /tɕʰjəŋ^{214}/; Z&L: /iŋ^{35}/, /tɕʰiŋ^{214}/), *yu3* and *yue4* (Lin: /ɥy^{214}/, /ɥe^{51}/; Z&L: /y^{214}/, /yɛ^{51}/), and *hun2* and *kuai4* (Lin: /xwən^{35}/, /kwai^{51}/; Z&L: /xuən^{35}/, /kuai^{51}/). They also differ according to certain vowels such as those found in the syllables *ou4*, and *hou4* (Lin: /ou^{51}/, /xou^{51}/; Z&L: /əu^{51}/, /xəu^{51}/), and *ye1*, *yan2*, and *juan3* (Lin: /je^{55}/, /jɛn^{35}/, /tɕɥɛn^{214}/; Z&L: /iɛ^{55}/, /ian^{35}/, /tɕyan^{214}/).

The fact that the two inventories differ should reminde us of Chao's nonuniqueness theory [84]. The uniqueness theory held that due to there being more than one way to represent a phonological system, there was no absolute better inventory per a given language, but rather an inventory more appropriate for a given purpose. Our purpose in creating a novel inventory was to ensure that a network built upon phonological similarity depended on a syllable inventory equally constructed on phonological similarity.

In the creation of the inventory, we did not seek to redefine Mandarin phonology through the classification of novel phonemes, but instead compare and contrast existing inventories with our participants' minimal pair creations so as to choose which phonemes best accounted for their productions. Thus, we first sought to identify where our participants' minimal pairs disagreed with the annotations of either the Lin and/or Z&L inventories. Agreement between the annotation systems and our participants' productions was assessed through the calculation of mean edit distance per stimuli. High agreement meant that a given stimuli's mean edit was near 1. Prior to calculating mean edit distance per stimuli, tonal neighbors were removed due to their segmental units being identical.

Stimuli of both high and low agreement were informative as to identifying changes in transcriptions that would follow our participants' minimal pair productions. For instance, the stimuli *an4*, which garnered a mean edit of 1.42 for both Z&L and Lin, garnered a lower mean edit of 1.26 for the newly formed Neergaard and Huang inventory (N&H). This was due to modifying the rime, /ɑŋ/, to /aŋ/ (N&H: /an^{51} ~ aŋ^{51}/, edit = 1; Lin and Z&L: /an^{51} ~ ɑŋ^{51}/, edit = 2). The mean edit for the stimuli, *qing3*, (Lin: 2.86; Z&L: 1.71; N&H: 1.71) illustrated that the addition of the glide, /j/, in

the Lin inventory created lower agreement (Lin: /tɕʰjəŋ^{214} ~ tɕʰin^{214}/, edit = 3; Z&L and N&H: /tɕʰiŋ^{214} ~tɕʰin^{214}/, edit = 1).

Another means to identify low agreement, and thus a means to improve the N&H inventory, was through targeting specific annotation choices. Lin's glide annotations /j,w,ɥ/ were shown to have lower agreement across multiple minimal pairs, such as *xin1~xian1* (Lin: /ɕin^{55} ~ ɕjɛn^{55}/, edit = 2; Z&L: /ɕin^{55} ~ ɕian^{55}/, edit = 1; N&H /ɕin^{55} ~ ɕiɛn^{55}/, edit = 1), *mo2~mu2* (Lin: /mwo^{35} ~mu^{35}/, edit = 2; Z&L: /mo^{35} ~ mu^{35}/, edit = 1; N&H /muo^{35} ~mu^{35}/, edit = 1), *quan2~qun2* (Lin: /tɕʰɥɛn^{35} ~tɕʰyn^{35}/, edit = 2; Z&L: /tɕʰyan^{35} ~tɕʰyn^{35}/, edit = 1; N&H /tɕʰyɛn^{35} ~tɕʰyn^{35}/, edit = 1). Similarly, both Lin and Z&L showed low agreement according to pinyin syllables that have the "ong" rime, annotated as /uŋ/. Participants preferred to produce phonological neighbors that contained /o/ rather than /u/, as can be seen in the example, *yong4~you4* (Lin: /juŋ^{51} ~jou^{51}/, edit = 2; Z&L: /iuŋ^{51} ~iəu^{51}/, edit = 2; N&H /ioŋ^{51} ~ioʊ^{35}/, edit = 1).

There were two cases in which the N&H inventory collapsed existing categories. Participants made neighbors ignoring the difference between the Lin and Z&L phonemes /ɤ/ and /ɘ/. By collapsing them into the single phoneme, /ɘ/, N&H reduced the mean edit compared to both Lin and Z&L for syllables such as *er2* (Lin: 2.5: Z&L: 2.5; N&H: 2.13), as is illustrated in the pair *er2~e2* (Lin: /ɤr^{35} ~ ɘ^{35}/, edit = 2; Z&L: /ɤɻ^{35}~ ɘ^{35}/, edit = 2; N&H /ɘr^{35} ~ ɘ^{35}/, edit = 1). N&H collapsed the Lin distinction of /ɔ/ and /ou/, and the Z&L distinction of /o/ and /əu/, into the single diphthong /oʊ/. This decision was based on garnering lower mean edits for the N&H inventory for syllables such as *bo1* (Lin: 1.7: Z&L: 2; N&H: 1.65), *mo2* (Lin: 1.94: Z&L: 1.94; N&H: 1.76), and *huo2* (Lin: 1.71: Z&L: 1.71; N&H: 1.59). It was also based on edit distances for minimal pairs such as *ou4~o1* (Lin: /ou^{51} ~ ɔ^{55}/, edit = 3; Z&L: /əu^{51} ~o^{55}/, edit = 3; N&H /oʊ^{51} ~oʊ^{35}/, edit = 1).

Examples of 10 syllables across the Lin, Z&L and N&H syllable inventories can be seen in Table 2. See Table 3 for the N&H phoneme inventory.

A final step in evaluating the three syllable inventories consisted of an ANOVA between edit distance values (excluding tonal neighbors): Lin (M: 1.90; SD: 0.92); Z&L (M:1.72; SD: 0.81); N&H (M: 1.67; SD: 0.79). The main effect was significant ($F=43.46$; $p < 0.001$). Pair-wise comparisons showed that both the Z&L ($p < 0.001$) and N&H ($p < 0.001$) inventories outperformed the Lin inventory. No significant difference was found between the edit distance values of Z&L and N&H.

2.2.2. Edit Information. Edit distance (including tonal neighbors) was used to calculate similarity according to the Lin, Z&L and N&H syllable inventories. Single-segment edits made up between 61 and 67% of correct responses (Lin: 61%; Z&L: 65%; N&H: 67%) while two-segment edits comprised over 20% (Lin: 23%; Z&L: 24%; N&H: 24%), three-segment edits accounted for around 10% (Lin: 12%; Z&L: 9%; N&H: 8%), and four- and five-segment edits combined were roughly 3% of correct responses (Lin: 4%; Z&L: 2%; N&H: 2%).

Table 2: Comparisons between syllable inventories.

Pinyin	Lin	Z&L	N&H
e	/ɤ/	/ɤ/	/ə/
ai	/ai/	/ai/	/aɪ/
ei	/ei/	/ei/	/eɪ/
o	/ɔ/	/o/	/oʊ/
ou	/ou/	/əu/	/oʊ/
ao	/ɑu/	/ɑu/	/aʊ/
ang	/ɑŋ/	/ɑŋ/	/aŋ/
yu	/ɥy/	/y/	/y/
yue	/ɥe/	/yɛ/	/yɛ/
yuan	/ɥɛn/	/yan/	/yɛn/

The single-segment edits can be further described by addressing which segments within the fully segmental schema (C_G_V_X_T) were altered to make a minimal pair (edit location) and the edit type (addition, deletion, or substitution) that was made per manipulation. The predominant segment to be changed within single-edit manipulations was that of lexical tone, which accounted for 34% of all correct responses across the three inventories. The second most often manipulated segment was the initial consonant, accounting for around 18% (Lin: 18%; Z&L: 18%; N&H: 19%). The remaining segments featured in less than 8% of all correct responses. The medial glide was manipulated roughly 2% across all inventories. The monophthong was around 5% (Lin: 4%; Z&L: 5%; N&H: 5%) and the final X between 3 and 8% (Lin: 3%; Z&L: 6%; N&H: 8%). As for edit type, the majority of manipulations made for correct responses were made through substitution (Lin: 55%; Z&L: 55%; N&H: 56%). Edits made from the addition of a segment accounted for between 5 and 8% (Lin: 5%; Z&L: 7%; N&H: 8%), while deletion type edits accounted for roughly 2% of correct responses (Lin: 1%; Z&L: 3%; N&H: 3%).

2.3. Discussion. In this experiment, participant-elicited minimal pairs served in the creation of a novel syllable inventory as well as provide insight into awareness of the units within the Mandarin syllable. As it stands currently, the Lin inventory was outperformed by both Z&L and the newly created N&H inventories with no statistical difference between the latter two. In terms of segmentation, while results show that all units are subject to manipulation, there was a strong prevalence towards two principle units: the unsegmented syllable and lexical tone. These results perhaps do not in themselves lessen the status of each segment but emphasize a tonal route for mental search of minimal pairs. Of another note on this experiment's findings is the fact that our Mandarin-speaking participants produced a lower percentage of single-edit responses (Lin: 61%; Z&L: 64%; N&H: 67%) than did the English speaking participants of [71] at 71%, [2] at 74.5%, and [73] at 84.21%. It is likely safe to assume that the lower values for the Luce and Large [71] study were the result of their participants having given spoken responses, whereas in both studies by Vitevitch and colleagues [2, 73] the recorded responses were written. Another reason for a lower percent

Table 3: N&H phoneme inventory.

	IPA	Sampa	Pinyin word	Sampa word	Ortho word		IPA	Sampa	Pinyin word	Sampa word	Ortho word
Vowels	a	a	b**a**3	p**a**3	把	Plosives	p	p	**b**u4	**p**u4	不
	ə	@	sh**e**4	S**@**4	蛇		p^h	P	**p**ao3	**P**aU3	跑
	e	e	g**e**i3	k**e**I3	给		k	k	**g**e0	**k**@0	个
	ɛ	E	y**e**3	i**E**3	也		k^h	K	**k**e4	**K**@4	课
	ɨ	1	zh**i**1	Z**1**1	之		t	t	**d**ou1	**t**oU1	都
	i	i	d**i**4	t**i**4	第		t^h	T	**t**a1	**T**a1	他
	ɪ	I	su**i**4	sue**I**4	岁	Fricatives	s	s	**s**uo3	**s**uo3	所
	o	o	ru**o**4	ru**o**4	若		f	f	**f**ang4	**f**aN4	放
	ʊ	U	cho**u**3	Co**U**3	丑		x	x	**h**ui4	**x**ueI4	会
	u	u	**w**o3	**u**o3	我	Affricates	ʂ	S	**sh**i4	**S**14	是
	y	y	**yu**an2	**y**En2	元		ɕ	X	**x**ia4	**X**ia4	下
Nasals	m	m	m**a**1	m**a**1	妈		tɕ	J	**j**iu4	**J**ioU4	就
	n	n	**n**eng2	**n**@N2	能		$tɕ^h$	Q	**q**ing3	**Q**iN3	请
	ŋ	N	xia**ng**3	Xia**N**3	想		ts^h	c	**c**ong2	**c**oN2	从
Liquids	l	l	**l**ie4	**l**iE4	列		$tʂ^h$	C	**ch**u1	**C**u1	出
	r	r	**r**ang4	**r**aN4	让		ts	z	**z**i4	**z**14	字
							tʂ	Z	**zh**e	**Z**@4	这

of single-edit manipulations might be due to the nature of our example pairs. Given our participants did produce examples of manipulations at all units, we decided in a second phonological association task to validate the performance of the three annotation systems using an explicit single-edit example, as was done in [71]. We expected this to increase the number of single-edit manipulations and in turn aid in discriminating which of the three inventories best aligns with our participants' manipulations.

3. Experiment 2

3.1. Methods

3.1.1. Participants. Of the thirty-four newly recruited native Mandarin speakers, one participant was removed from further consideration because they rated Mandarin as being their nondominant language. The thirty-three remaining participants reported native-level fluency in Mandarin (Female: 22; Age: M, 23; SD, 4). None of the participants reported speech, hearing, or visual disorders. Participant characteristics did not differ from those from the first experiment in self-rated spoken English proficiency (M: 6.55; SD: 1.23), traditionally Mandarin-speaking region (Guanhua: 23; non-Guanhua: 10), or number of Chinese languages/dialects spoken (M: 2.39; SD: 0.74).

As with Experiment 1, all participants gave their informed consent and were compensated with 50HKD for their participation as stipulated by the local ethics committee.

3.1.2. Stimuli. The stimuli for Experiment 2 consisted of 200 test items and 10 practice items. Two items, however, were discounted for not existing in the www.zdic.net dictionary, reducing our total to 198 stimuli test items. Ninety-seven stimuli were used from the Experiment 1 stimuli set. The 101 new stimuli were created with the same voice and procedure. A determining factor in the selection of new stimuli and rejection of stimuli from Experiment 1 was whether or not lexical frequency could be accounted for using the word list from Subtlex-CH [85]. The Subtlex-CH wordlist, created through aggregated movie subtitles, was chosen because the subtitle genre has been shown to greater predict language processing tasks when compared to frequencies calculated from written sources [85, 86]. Further information on the transcription of the wordlist's 99,125 Chinese characters to pinyin and subsequent sampa can be found in the Database of Mandarin Neighborhood Statistics [87].

As can be seen in Table 4, we included each of the base rime syllables accompanied by between three to six consonant neighbors. As with the stimuli in Experiment 1, certain stimuli lacked base rimes, while others did not have consonant neighbors. Those stimuli with only three consonant neighbors were tied to the base rime syllables *yuan2* and *yong3*. They were limited to three consonant neighbors because there are only the three possible onsets, "*j*, *q*, *x*" /tɕ, $tɕ^h$, ɕ/, available for these base rimes and we did not want to repeat a nontonal syllable.

3.1.3. Procedure. The procedure differed from Experiment 1 in that no pretraining phase was given. In place of this, participants were given oral instructions as to what consisted of a similar sounding monosyllable through the use of the target syllable ling2 (e.g., 零), which they were told had the neighbors: ling4 (e.g., 另), ning4 (e.g., 宁), lang2 (e.g., 狼), and lin2 (e.g., 磷). All other procedural aspects were the same as in Experiment 1.

3.2. Results. As with Experiment 1, the same transcription procedure was followed. Missing (53), identical (210), nonitem (505), and semantically related responses (3) were excluded from the analysis.

Table 4: Experiment 2 test stimuli.

Base Rime	+C	+C	+C	+C	+C	+C
a1	*ba3*	*fa3*	*ka3*	*la1*	*ma1*	
ai4	*cai2*	*dai4*	*gai1*	*mai3*	*zai3*	
an4	*ban1*	*nan2*	*ran2*	*san3*	*shan1*	*zhan4*
ang2	*gang1*	*rang4*	*sang1*	*shang4*	*tang3*	
ao1	*gao3*	*lao3*	*mao1*	*pao4*	*zao3*	
e4	*che1*	*de2*	*ge1*	*ke3*	*zhe4*	
ei4	*bei3*	*hei1*	*fei2*	*mei2*	*pei2*	
en1	*fen4*	*hen3*	*men2*	*ren2*	*zhen4*	
	feng1	*reng1*	*sheng3*	*zeng4*	*zheng4*	
er3						
	ci4	*chi1*	*ri4*	*shi2*	*si3*	*zi3*
	hong2	*cong2*	*nong4*	*song4*	*zhong1*	*zong3*
ou4	*bo1*	*fo2*	*hou4*	*mo2*	*rou4*	
wa1	*gua4*	*hua2*	*kua3*	*shua1*	*zhua1*	
wai4	*guai4*	*huai4*	*kuai4*	*shuai4*		
wan2	*guan3*	*huan4*	*ruan3*	*suan4*	*tuan2*	
wang4	*guang1*	*huang2*	*shuang3*	*kuang2*	*zhuang1*	
wei2	*chui1*	*sui4*	*shui3*	*tui3*	*zui4*	
wen4	*hun4*	*lun2*	*gun3*	*zun1*		
weng4						
wo4	*cuo4*	*duo1*	*huo2*	*ruo4*	*shuo1*	
wu3	*chu1*	*du4*	*fu4*	*ru2*	*zhu4*	
ya4	*dia3*	*jia1*	*lia3*	*qia1*	*xia4*	
yan3	*bian3*	*mian2*	*pian4*	*qian2*	*tian2*	
yang3	*jiang1*	*liang2*	*niang2*	*qiang2*	*xiang3*	
yao4	*diao4*	*miao2*	*piao4*	*tiao1*	*xiao3*	
ye2	*die1*	*jie1*	*bie2*	*lie4*	*xie2*	
yi1	*ji1*	*li3*	*ni3*	*qi3*	*di1*	
yin1	*jin4*	*qin2*	*xin1*	*pin1*	*min2*	
ying2	*bing1*	*jing3*	*qing3*	*ting1*	*ming2*	
yong3	*jiong3*	*qiong2*	*xiong2*			
you3	*diu1*	*jiu3*	*liu1*	*niu2*	*qiu2*	
yu3	*ju2*	*lv4*	*nv3*	*qu4*	*xu1*	
yuan2	*juan4*	*quan2*	*xuan3*			
yue4	*jue2*	*nue4*	*que1*	*xue2*		
yun4	*jun1*	*kun4*	*qun2*	*xun1*		

We again evaluated which of the three syllable inventories optimally accounted for phonological similarity according to our participants' minimal pair productions. Repeating the same procedure, we excluded tonal neighbors prior to conducting an ANOVA on the edit distances of the three inventories: Lin (M: 1.86; SD: 0.87); Z&L (M: 1.71; SD: 0.78); N&H (M: 1.64; SD: 0.75). The main effect was significant ($F=65.3$; $p < 0.001$). Pair-wise comparisons showed that both the Z&L ($p < 0.001$) and N&H ($p < 0.001$) inventories outperformed the Lin inventory. Meanwhile the N&H inventory outperformed the Z&L inventory ($p = 0.002$).

Edit information, including edit distance, location, and type, was then calculated for the three inventories. All calculations were derived from the correct responses, including tonal neighbors.

Single-segment edits accounted for between 68 and 73% (Lin: 68%; Z&L: 71%; N&H: 73%). Two-segment edits accounted for around 21% (Lin: 20%; Z&L: 22%; N&H: 20%). Three-segment edits accounted for 5 to 10% (Lin: 10%; Z&L: 7%; N&H: 5%), and four- and five-segment edits combined were between 1 and 2% (Lin: 2%; Z&L: 1%; N&H: 1%).

Edit location for the single-segment edits again was dominantly at the lexical tone position, accounting for 46% of correct responses. The second most common manipulation, at 15%, again occurred at the initial consonant. The remaining syllable position saw a combined 5 to 16% instance of manipulation (Final X: Lin: 3%; Z&L: 5%; N&H: 7%, monophthong: Lin: 3%; Z&L: 4%; N&H: 4%, and medial glide: 1%).

Edit type again was dominantly substitution, occurring between 64 and 66% (Lin: 64%; Z&L: 65%; N&H: 66%). Edits

made from the addition of a segment accounted for between 3 and 5% (Lin: 3%; Z&L: 4%; N&H: 5%), while deletion type edits accounted for roughly 2% of correct responses (Lin: 1%; Z&L: 2%; N&H: 2%).

3.3. Discussion. The second phonological association task identified an optimal annotation system while providing repeated evidence of segmentation biases, specifically towards the manipulation of lexical tone while maintaining a whole syllable. Changes made in comparison to Experiment 1 included (1) changing instructions so as to provide a single-edit example and (2) increasing the number of stimuli, which respectively increased the percentage of single-edit productions (Experiment 1: 61-65%; Experiment 2: 67-73%) and gave greater discriminative power in identifying the newly formed N&H inventory as the optimal syllable inventory. In applying the principle of the nonuniqueness theory [84], we can surmise that the N&H inventory, built on phonological similarity, is the optimal choice to model Mandarin vocabulary in a phonological network that is as well constructed on phonological similarity.

4. Phonological Segmentation Neighborhoods

The goal of previous investigations into phonological networks has been to infer aspects of the nature of language processing and/or the development of the lexicon from constructed, random, and real language graphs. A number of topological measures have been used. Those that we will be reporting on come from the same six studies [9, 11–15]. The igraph package in R [88] was used for the construction and measurement of all the following graphs.

The first value to consider is degree. When expressed at the word level, annotated as k, it is the number of single-edit neighbors a given word has. At the topological level, annotated as $\overline{k}$, it is the mean of neighbors per node across the entire network, or from the network's largest fully connected subgraph, also referred to as the network's giant component. The giant components of phonological networks studied thus far have been shown to take between 32-66% of available nodes, which is lower than phonological networks built from artificial corpora [9]. All topological measures featured below will be reported from each network's giant component.

Interconnectedness between neighbors is expressed through the measure known as clustering coefficient. At the word level, annotated as CC, it is the proportion of neighbors who are also neighbors of each other. The mean value taken at the macrolevel is annotated as $\overline{CC}$. Phonological networks have shown $\overline{CC}$ values of between 0.191-0.383 for giant components.

Another measure of the relationship to the density of interconnectedness is the correlation between the density of a given node and the density of its neighbors, known as mixing by degree (M) [89, 90]. When positive, referred to as assortative mixing by degree, the value indicates that the network's nodes tend to have dense nodes connected with other dense nodes. Thus far, phonological networks, whether from real vocabulary lists or artificially constructed vocabularies, have all been assortative. Networks constructed from real vocabulary lists have shown M values between 0.556 and 0.762.

A final measure of network density, which we annotate as $\overline{L}$, is that of a networks' mean shortest path length. It is the average distance between a given node and the rest of the nodes within the giant component and thus a measure of spreading through the network. Phonological networks have been shown to have $\overline{L}$ values between 6.08 and 10.40 for their giant components.

We categorized the PSNs as to whether they had small world characteristics. A network that shows small world characteristics ($\overline{CC} > \overline{CC}\text{-}RN$; $\overline{L} > \overline{L}\text{-}RN$) has values of $\overline{CC}$ and $\overline{L}$ greater than those generated from random networks ($\overline{CC}\text{-}RN$, $\overline{L}\text{-}RN$). We report on the mean and standard deviations of 10 iterations of Erdos-Renyi random networks constructed from the same number of nodes and edges as the networks they were compared to. The small world structure is believed to aid speed during search [91] and thus generalize to the spreading of lexical activation [13, 15]. All language networks thus far have shown small world characteristics in their giant components.

Finally, we also report on whether the PSNs' degree distributions can be described as having a power-law degree distribution. Vitevitch [13] drew attention to the distributions of phonological networks as a possible cue to vocabulary formation due to the association between power-law degree distributions with self-organization [92] and the two principles underlying scale-free networks: growth and preferential attachment [93, 94]. Preferential attachment describes a process whereby new nodes establish connections to already densely connected nodes. A limitation to this association of scale-free characteristics and power-law degree distributions is that scale-free networks can come about from other growth methods [95]. Phonological networks have not shown clear cut power-law distributions, but instead a power-law with cut-off [3, 5, 9, 14, 15]. The term cut-off refers to the process of choosing a starting point from which the distribution is fit [96], meaning that only a portion of a given distribution is being described. The present degree distributions were fitted using [97].

While the initial studies were optimistic about which of the many variables were indicative of cognitive processes [13, 15], few seem to be likely candidates. The construction of pseudo lexicons and their subsequent comparisons to real phonological networks has shown that small world characteristics are not intrinsic to the nature of vocabulary [11, 98] and that preferential attachment is not a likely account of vocabulary growth due to portions of power-laws also occurring from distributions made through random sampling [9]. Turnbull and Peperkamp [12] placed their hope in assortative mixing by degree due to it being the only value to distinguish an English phonological network from 5 types of random graphs. Stella and Brede [9] similarly had higher assortativity for their English network when compared to their constructed networks. Yet, it is not clear whether a difference of either 0.103 [12] or 0.117 [9] between their real networks and their second highest constructed networks is meaningful. Other studies have suggested that word length

Table 5: Mandarin segmentation schemas according to the example monosyllables, *lian2* /liɛn^{35}/ and *liao3* /liaʊ^{214}/.

Without Tone			With Tone		
C_V_C	/l_iɛ_n/	/l_iaʊ/	C_V_C_T	/l_iɛ_n_35/	/l_iaʊ_214/
C_G_V_C	/l_i_ɛ_n/	/l_i_aʊ/	C_G_V_C_T	/l_i_ɛ_n_35/	/l_i_aʊ_214/
C_G_V_X	/l_i_ɛ_n/	/l_i_a_ʊ/	C_G_V_X_T	/l_i_ɛ_n_35/	/l_i_a_ʊ_214/
C_G_VX	/l_i_ɛn/	/l_i_aʊ/	C_G_VX_T	/l_i_ɛn_35/	/l_i_aʊ_214/
C_GVX	/l_iɛn/	/l_iaʊ/	C_GVX_T	/l_iɛn_35/	/l_iaʊ_214/
CG_V_X	/li_ɛ_n/	/li_aʊ/	CG_V_X_T	/li_ɛ_n_35/	/li_aʊ_214/
CG_VX	/li_ɛn/	/li_aʊ/	CG_VX_T	/li_ɛn_35/	/li_aʊ_214/
CGVX	/liɛn/	/liaʊ/	CGVX_T	/liɛn_35/	/liaʊ_214/

plays a unique role in the networks. Network statistics are influenced by word length [9, 12, 98], because of the negative correlation found between length and phonological similarity according to the single-edit metric. Languages with greater morphological richness have shown sparser distributions [14, 98], hinting at cross-linguistic differences based on graph measures.

4.1. Constructing the PSNs. With the N&H inventory validated, it was then used to create a database of neighborhood statistics from all schematic representations proposed or suggested. In order to provide all possible permutations we add the segmented diphthongal schema, previously proposed for Taiwanese speakers [99] in its nontonal (C_G_V_C) and tonal form (C_G_V_C_T). The possibility of diphthongs was proposed for Mandarin by [100]. Table 5 presents sixteen segmentation schemas, each with two example syllables.

Lexical frequencies and subsequent neighborhood frequency counts (the average frequency of a words' neighbors) were again adapted from Subtlex-CH [85] as detailed in the Database of Mandarin Neighborhood Statistics [87]. Prior to calculating phonological neighbors, all homophones were collapsed into single items, and their frequencies summed, i.e., the definition of a phonological word. Each PSN was then created from the top 30,000 most frequent phonological words, roughly the same size as the Mandarin network analyzed by Arbesman et al. [14]. This led to slight differences in degree (PND) from the existing resource of similar structure and content [87] that calculated similarity from the top 17,000 most frequent phonological words. Monosyllables that were featured in the stimuli but that were not present in the Subtlex-CH word list were added for the sake of calculating their degree, but were given a frequency count of 1 and thus were not part of the top 30,000 phonological words from which degree calculations were made. Lastly, we removed edges between monosyllables in the CGVX PSN, consisting of 397 monosyllabic neighbors per target monosyllable word, due to there being no meaningful relationship between them.

4.2. Topology. As can be seen in Table 6, the PSNs exhibit network characteristics both within and outside expected ranges compared to past phonological networks. Unlike previous networks, $\overline{L}$ (2.79-17.72), M (0.454-0.918), and the proportion of the network covered by the giant component (Size: 30.53-88.15%) all showed a large range of values, some of which were double those previously reported. $\overline{CC}$ (0.247-0.628) was perhaps the only measure with relative stability across PSNs. In line with past networks, all PSNs exhibited small world characteristics ($\overline{CC} > \overline{CC}$-$RN$; $\overline{L} > \overline{L}$-$RN$).

In Table 6, we characterize the sixteen Mandarin PSNs according the number of units within each PSNs' maximal syllable (Units). In Figure 1, we see that both Units and lexical tone determine how each PSN patterns according to their network characteristics. What first stands out is the distance the unsegmented PSNs (CGVX, CGVX_T) take from their segmented counterparts. While CGVX_T groups according to the nontonal PSNs in Size (a) and $\overline{L}$ (b), it stands apart in M (c), yet is similar to CGVX in its high $\overline{CC}$ (d). CGVX, meanwhile, has a uniquely high $\overline{k}$ (139.97) and Size, similar to the collaboration networks reported by Newman [101]. $\overline{L}$ in contrast is very low, illustrating that high $\overline{k}$ and Size equate short distances between any given neighbors. Only in M does CGVX pattern according to the nontonal PSNs. The segmented PSNs on the other hand show some gradient distributions. Size shows a negative trend for greater segmentation, particularly for tonal PSNs, which is opposite to the positive trend found in $\overline{L}$. There is no linear effect of segmentation for either M or $\overline{CC}$. M exhibits the only split distribution among the network statistics. Unfortunately, there is no immediate indication why the low M group (C_GVX_T, C_V_C_T, CG_V_X_T) would have roughly half of the values of the high M group (CG_VX_T, C_G_VX_T, C_G_V_C_T, C_G_V_X_T).

In Table 6 we see that not all PSNs contain portions of power-law distributions. Those that did contain power-law portions were both nontonal and tonal and of varying unit lengths (nontonal: CGVX, C_GVX, C_V_C, CG_V_X; tonal: CGVX_T, CG_VX_T, CG_V_X_T, C_G_VX_T, C_G_V_C_T, C_G_V_X_T), which similarly can be said for those that did not (nontonal: CG_VX, C_G_VX, C_G_V_C, C_G_V_X; tonal: C_GVX_T, C_V_C_T). Segmental units were also not to blame seeing as all individual units and their collapsed combinations occur in both distribution groups.

4.3. Syllable Length. Of the top 30,000 phonological words, monosyllables account for 3.80% (n=1,141), disyllables 72.17% (n=21,652), trisyllables 14.84% (n=4453), quadrasyllables 8.73% (n=2618), and less than 1% for the remaining 5-, 6-, and 7-syllable phonological words (n=136). In Figure 2 we

Table 6: Topological network measures of the PSNs' giant components.

	CGVX	C_GVX	CG_VX	C_V_C	CG_V_X	C_G_VX	C_G_V_C	C_G_V_X
Units	1	2	2	3	3	3	4	4
Size	88.15	71.91	70.41	70.52	69.1	70.33	69.13	68.79
$\overline{k}$	139.97	14.53	13.2	10.19	10.54	11.73	9.79	8.93
M	0.577	0.602	0.628	0.613	0.689	0.593	0.622	0.633
$\overline{L}$	2.79	5.31	5.58	6.49	7.47	5.77	6.85	7.65
$\overline{L}$-RN	2.47 (1.93^{e-5})	4.01 (2.75^{e-4})	4.14 (4.09^{e-4})	4.56 (3.16^{e-3})	4.50 (3.16^{e-3})	4.33 (6.59^{e-4})	4.62 (1.13^{e-3})	4.79 (1.14^{e-3})
$\overline{CC}$	0.578	0.319	0.336	0.303	0.435	0.278	0.310	0.340
$\overline{CC}$-RN	5.29^{e-3} (6.25^{e-6})	6.53^{e-4} (4.01^{e-5})	6.10^{e-4} (3.30^{e-5})	4.81^{e-4} (4.19^{e-5})	4.97^{e-4} (3.48^{e-5})	5.36^{e-4} (3.19^{e-5})	4.54^{e-4} (4.86^{e-5})	4.25^{e-4} (3.24^{e-5})
Cut-off/p	226/∗∗	27/∗	24/*NS*	19/∗∗	20/∗∗	37/*NS*	31/*NS*	27/*NS*
	+T	+T	+T	+T	+T	+T	+T	+T
Units	2	3	3	4	4	4	5	5
Size	71.89	49.11	48.88	34.39	31.02	45.1	35.29	30.53
$\overline{k}$	25.64	3.61	4.98	3.10	3.14	4.61	4.47	4.51
M	0.733	0.538	0.918	0.454	0.470	0.894	0.891	0.900
$\overline{L}$	5.40	12.12	12.68	15.15	17.47	12.44	14.74	17.72
$\overline{L}$-RN	3.42 (1.81^{e-4})	7.57 (1.47^{e-2})	6.16 (3.20^{e-3})	8.17 (2.48^{e-2})	8 (2.23^{e-2})	6.40 (7.84^{e-3})	6.37 (9.66^{e-3})	6.24 (8.22^{e-3})
$\overline{CC}$	0.628	0.303	0.460	0.247	0.335	0.358	0.388	0.416
$\overline{CC}$-RN	1.21^{e-3} (1.90^{e-5})	2.16^{e-4} (9.25^{e-5})	3.27^{e-4} (5.98^{e-5})	2.99^{e-4} (1.47^{e-4})	3.85^{e-4} (9.07^{e-5})	3.43^{e-4} (8.30^{e-5})	3.96^{e-4} (9.51^{e-5})	4.91^{e-4} (1.19^{e-4})
Cut-off/p	71/∗∗	10/*NS*	3/∗∗	10/*NS*	6/∗∗	4/∗∗	3/∗∗	3/∗∗

Note: Units = number of units within each PSN's maximal syllable; Size = percent of nodes covered by the giant component; $\overline{k}$ = mean degree; M = mixing by degree; $\overline{L}$ = mean shortest path length; $\overline{L}$-RN = $\overline{L}$ of 10 iterations of random networks (mean standard deviation of $\overline{L}$-RN); $\overline{CC}$ = mean clustering coefficient; $\overline{CC}$-RN = $\overline{CC}$ of 10 iterations of random networks (mean standard deviation of $\overline{CC}$-RN); Cut-off/p = properties of a power-law degree distribution, such that Cut-off denotes where in the distribution the calculation begins, and p, the probability of said distribution accounting for a power-law distribution, expressed as either NS (nonsignificant), ∗ ($p < 0.05$), or ∗∗ ($p < 0.01$)

illustrate the distributions for $\overline{k}$ and $\overline{CC}$, for monosyllables and disyllables, according to the number of maximal syllable units (Units) in each PSN. Because of the difference between segmented and unsegmented PSNs, we will consider them separately.

In Figures 2(a) and 2(b) we see that greater segmentation and the addition of lexical tone led to fewer neighbors for both monosyllables and disyllables. The PSNs with the lowest $\overline{k}$ were built from five units and are both tonal (C_G_V_C_T, C_G_V_X_T). Conversely, the segmented syllables that have only two units, CG_VX and C_GVX, are both nontonal and have the highest $\overline{k}$ of the segmented PSNs. Compared to $\overline{k}$, the story of $\overline{CC}$ for monosyllables and disyllables is less clear. There is no trend between Units and $\overline{CC}$ for monosyllables (Figure 2(c)) from segmented PSNs. Conversely, $\overline{CC}$ among disyllables of segmented PSNs are affected by lexical tone. Figure 2(d) shows that nontonal PSNs all have higher $\overline{CC}$ than tonal PSNs.

To account for the outlier behavior of unsegmented PSNs, we first address the switch in $\overline{k}$ between monosyllables ($\overline{k}$ = 286) and disyllables ($\overline{k}$ = 26). For monosyllables in CGVX_T, every phonological word of a given tone assignment is a neighbor of every other monosyllable with that same tone, leading to 5 distinct subgraphs (tones 0-4). The complete interconnectedness of neighbors for monosyllables means that these words have CC values nearing 1, as seen in Figure 3(c). Disyllabic words of the CGVX_T PSN, on the other hand, have 3 of 4 units that must match with another word to classify as a neighbor and as such do not diverge greatly from the linear relationship between $\overline{k}$ and Units of the segmented PSNs. The increase in Units reduces not just $\overline{k}$, but also $\overline{CC}$.

In contrast to the CGVX_T PSN, the nontonal unsegmented PSN (CGVX) has an opposite switch in $\overline{k}$ for monosyllables ($\overline{k}$ = 117) and disyllables ($\overline{k}$ = 186). The number of neighbors increases for disyllabic words due to the ability of nontonal disyllabic words to link to monosyllables, other disyllables, and trisyllables. Unlike with monosyllables, in which $\overline{k}$ and $\overline{CC}$ distributions differ according to Units, disyllables show a linear relation between Units and both $\overline{k}$ (b) and $\overline{CC}$ (d).

To aid in comprehending the role of segmentation and lexical tone on network features at the word level, in Figure 3 we illustrate the tonal monosyllabic word *niao3* /niaʊ^{214}/ and its nontonal counterpart *niao* /niaʊ/. The nontonal *niao* of CGVX (Figure 3(a)) is the sole monosyllable in

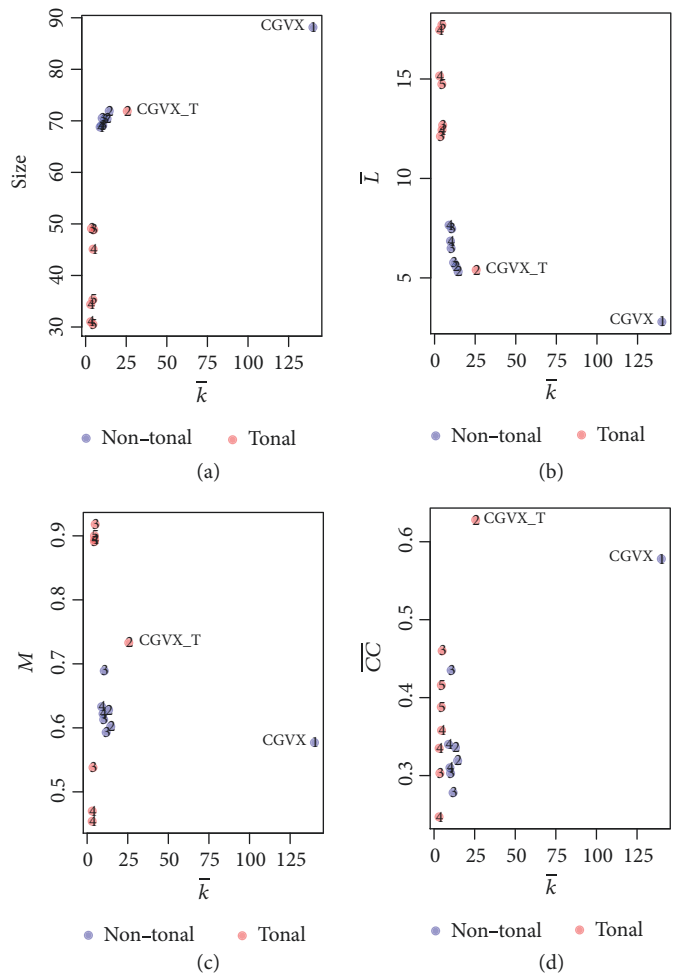


Figure 1: Topological features of each PSN according to $\overline{k}$ (mean degree) and (a) Size (percent of giant component), (b) $\overline{L}$ (mean shortest path length), (c) M (mixing by degree), and (d) $\overline{CC}$ (mean clustering coefficient). Units (the number of units within each PSNs' maximal segmentation schema) is featured within each data point.

a network of disyllables. Through the addition of lexical tone (Figure 3(b)), all disyllables are excluded, and neighbor classification is based on whether the monosyllables share the same tone. Meanwhile, a segmented *niao* (Figure 3(c)) has both monosyllabic neighbors that differ by a single segment, and disyllabic neighbors, such as *ni hao* /ni xao/. This is not the case for *niao3* (Figure 3(d)), which has only other monosyllables as neighbors.

4.4. Discussion. Sixteen Mandarin PSNs were constructed that differed according to both syllable segmentation and lexical tone. Network statistics revealed that both characteristics determined what constituted similarity between phonological words. Greater segmentation and the presence of tone mean less density for segmented neighbors. Plots of the PSNs' $\overline{k}$ was informative as to each segmented network's Size (larger for nontonal PSNs), $\overline{L}$ (larger for tonal PSNs), M (assortative for all PSNs, and split for tonal PSNs), and $\overline{CC}$ (no clear trend). Unsegmented PSNs, in contrast, behaved differently from segmented PSNs for each network measure at both the scale of the giant component and when isolating monosyllables and disyllables. Inspection of word-level graphs illustrated that for monosyllables, the presence of tone limited the choice of available neighbors to other monosyllables, while monosyllables of nontonal PSNs had both monosyllabic and disyllabic neighbors. For unsegmented PSNs, this was exacerbated, such that monosyllabic words from the nontonal unsegmented PSN (CGVX) had only disyllabic neighbors, and monosyllabic words from the tonal unsegmented PSN (CGVX_T) had only monosyllabic tonal neighbors.

We now turn to the principle goal of inspecting topological features of language networks, and phonological networks in particular. Under the conceit that properties of language processing and vocabulary formation can be inferred from phonological networks built from vocabulary lists, we ask whether we can predict which of the sixteen PSNs is the most likely candidate for Mandarin based on previous network measures.

Previous phonological networks showed between 32 and 66% in Size. This range includes all segmented and tonal PSNs, while excluding the nontonal segmented group and

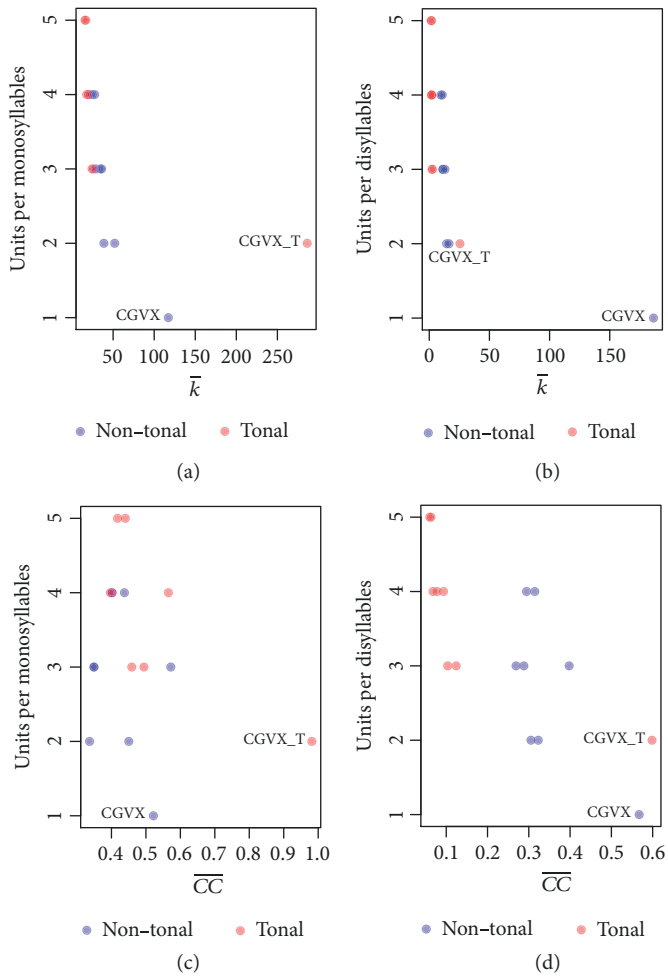


Figure 2: Units (the number of units within each PSNs' maximal segmentation schema) plotted against $\overline{k}$ (mean degree) for monosyllables (a) and disyllables (b), and $\overline{CC}$ (mean clustering coefficient) for monosyllables (c) and disyllables (d).

both unsegmented PSNs, which fell within a range of 69-88%. Thus, using Size alone would predict the likely candidate as both tonal and segmented.

Phonological networks have shown between 0.191 and 0.383 in mean clustering coefficient ($\overline{CC}$). The tonal and nontonal segmented PSNs comprise one group falling in between 0.247 and 0.460. Using $\overline{CC}$ as an indicator would exclude the unsegmented PSNs that have higher values (CGVX: 0.578; CGVX_T: 0.628).

Despite the possibility of a phonological network falling within the negative range (disassortative) of M, phonological networks have been positive (assortative), falling between 0.556-0.762. Our PSNs were also assortative, but did not follow a specific trend. Nontonal PSNs were tightly grouped between 0.577-0.689, which patterned similarly with previous phonological networks. The split in distributions for tonal PSNs meant that the low group fell below (C_V_C_T: 0.454; CG_VX_T: 0.470) the expected range, while the high group far above (CG_VX_T: 0.918; C_G_V_X_T: 0.900; C_G_VX_T: 0.894; C_G_V_C_T: 0.891). Only two tonal networks were near or within the expected range (C_GVX_T: 0.538; CGVX_T: 0.733).

Phonological networks have shown values in $\overline{L}$ between 6.08 and 10.40. This range excludes the nontonal unsegmented PSN (CGVX: 2.79), and the tonal segmented PSNs, which had higher values falling between 12.12 and 17.72. The past networks however are near or within the range of the nontonal segmented PSNs (5.31-7.65) and the tonal unsegmented PSN (CGVX_T: 5.40).

Finally, while all PSNs met the conditions for small world networks, not all were suggestive of being scale-free networks. Neither segmentation nor tone accounted for why four nontonal networks (CG_VX, C_G_VX, C_G_V_C, C_G_V_X) and two tonal networks (C_GVX_T, C_V_C_T) did not have power-law degree distributions.

No single PSN patterned according to past phonological networks. However, discounting Size, three nontonal segmented PSNs, C_GVX, C_V_C,and CG_V_X, meet the remaining criteria. In the next section, we evaluate the reaction times from Experiment 2 with the goal of identifying which of the sixteen PSNs was the likely candidate.

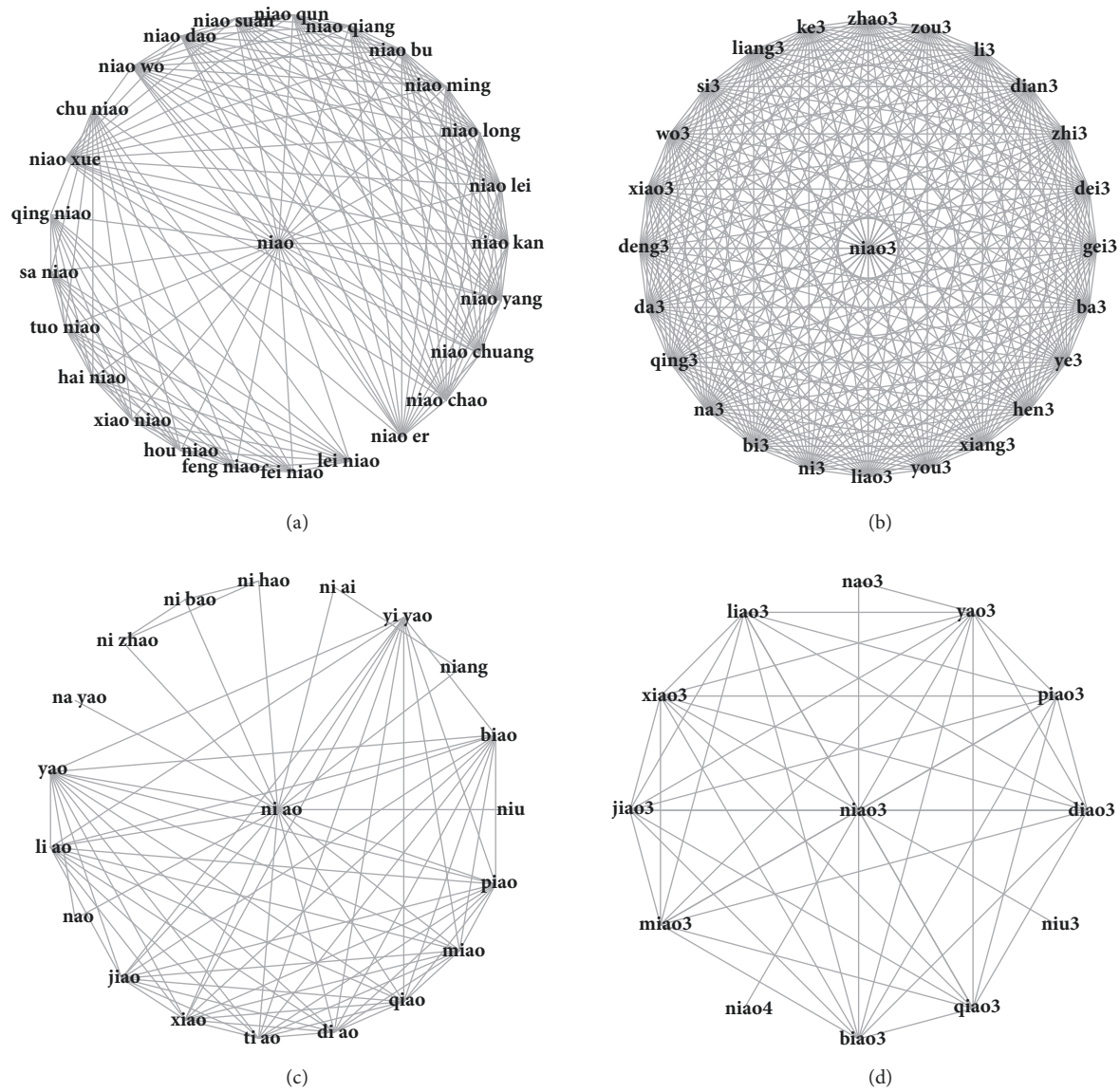


FIGURE 3: Word-level phonological networks for the monosyllabic word *niao3* /niau^{214}/, according to (a) the nontonal unsegmented PSN (CGVX), (b) the tonal unsegmented PSN (CGVX_T), (c) the nontonal fully segmented PSN (C_G_V_X), and (d) the tonal fully segmented PSN (C_G_V_X_T). For visualization purposes, (b) was restricted to just 20 visible neighbors due to all CGVX_T values being well over 200.

5. Model Selection Procedure

The goal of the current methods was to identify an optimal PSN through the lexical statistics that were tied to them. From the outset, this implied the identification of an optimal model in comparison to many other models. We began with backwards selection to identify which of the participant-related and stimuli-related predictors merited inclusion in the random effects structure. The purpose of having a complex random effects structure was not to increase generalizability of a confirmatory analysis, as proposed by Barr et al. [102], but instead to both restrict the current exploratory analysis from overestimating the effects of our network predictors and to guide future confirmatory analyses in dealing with participant- and stimuli-related characteristics. As is a current norm in the psycholinguistic literature, Subject and Item were included as random intercepts in all models featured. Upon identifying a random effects structure, sixteen full models were assessed according to R^2 using the Kenward-Roger approximation [103]. We used the "r2glmm" package in R [104] to (1) measure both marginal R^2 for full models, and semipartial R^2 for each fixed effect, and (2) to perform an R^2 difference test between our top ranked models.

Reaction times were measured offline using SayWhen [105]. One participant was excluded due to mean reaction times greater than 2.5 standard deviations above the group mean. Outliers with reaction times greater than 3000ms and lower than 415ms were then excluded, followed by three stimuli (*dia3*, *fo2*, *gun3*) with error rates greater than a third of the number of participants. From the remaining 6,240 trials, 31 false starts, 391 nonitems, 187 identical items, 24 missing,

Table 7: Experiment 2 lexical statistics.

	HD M(SD)	*k* M(SD)	*CC* M(SD)	Freq M(SD)	NF M(SD)
CGVX	18.91	119.86	0.53	4.15	2.47
	(14.76)	(85.67)	(0.11)	(0.89)	(0.47)
C_GVX	18.91	51.18	0.34	4.15	4.91
	(14.76)	(13.66)	(0.1)	(0.89)	(0.43)
CG_VX	18.93	37.13	0.46	4.15	4.8
	(14.76)	(11.23)	(0.16)	(0.89)	(0.34)
C_V_C	18.91	33.48	0.34	4.15	4.91
	(14.76)	(9.67)	(0.1)	(0.89)	(0.5)
CG_V_X	18.93	28.13	0.57	4.15	4.85
	(14.76)	(9.58)	(0.19)	(0.89)	(0.4)
C_G_VX	19.02	34.67	0.33	4.15	4.85
	(14.77)	(9.34)	(0.08)	(0.89)	(0.41)
C_G_V_C	19.02	26.41	0.37	4.15	4.89
	(14.77)	(7.24)	(0.12)	(0.89)	(0.45)
C_G_V_X	19.02	23.71	0.4	4.15	4.9
	(14.77)	(6.82)	(0.14)	(0.89)	(0.45)
CGVX_T	5.58	285.87	0.99	3.71	4.18
	(4.59)	(34.04)	(0.01)	(0.99)	(0.18)
C_GVX_T	5.58	24.98	0.49	3.71	4.22
	(4.59)	(8.19)	(0.12)	(0.99)	(0.47)
CG_VX_T	5.58	22.86	0.51	3.71	4.17
	(4.59)	(8.13)	(0.14)	(0.99)	(0.49)
C_V_C_T	5.58	16.45	0.39	3.71	4.24
	(4.59)	(6.35)	(0.12)	(0.99)	(0.53)
CG_V_X_T	5.58	18.82	0.57	3.71	4.22
	(4.59)	(7.74)	(0.19)	(0.99)	(0.56)
C_G_VX_T	5.58	18.16	0.39	3.71	4.22
	(4.59)	(7.5)	(0.11)	(0.99)	(0.5)
C_G_V_C_T	5.58	15.26	0.39	3.71	4.27
	(4.59)	(6.06)	(0.14)	(0.99)	(0.54)
C_G_V_X_T	5.58	14.19	0.41	3.71	4.28
	(4.59)	(5.81)	(0.16)	(0.99)	(0.55)

and 1 semantically related item were excluded, giving us a mean of 1530ms (SD: 557ms).

After exclusion, participants' responses consisted of edit distances between 1-5: Edit 1, 3532 observations (M: 1496ms; SD: 556ms); Edit 2, 934 observations (M: 1617ms; SD: 550ms); Edit 3, 245 observations (M: 1642ms; SD: 553ms); Edit 4, 49 observations (M: 1766ms; SD; 552ms); Edit 5, 6 observations (M: 1756ms; SD: 677).

Age, sex, self-rated spoken English, and whether a speaker was from a traditionally Guanhua speaking region were all nonsignificant, as were segment length (SegLen 1: 5, SegLen 2: 47; SegLen 3: 98; SegLen 4: 45) and lexical tone (tone 1: 49; tone 2: 47; tone 3: 43; tone 4: 56). The number of Chinese languages/dialects spoken by our participants (Num_Chinese) did significantly account for a portion of the variance. Our preliminary models revealed that higher values of Num_Chinese led to slower reaction times. Due to this variable representing variation at the participant level, it was added to the random effects structure as a random slope of Subject.

The fixed effects under consideration include Edit and five variables that vary due to PSN construction: homophone density (HD), lexical frequency (Freq), neighborhood frequency (NF), word-level degree (*k*), and word-level clustering coefficient (*CC*). All mean and standard deviations for the 80 network predictors (16 PSNs $*$ 5 network predictors) can be found in Table 7. Edit was not centered due to it being an interval measurement of only 5 levels, while the variables representing the PSNs (HD, *k*, *CC*, Freq, NF) were centered. Model selection output can be seen in Table 8.

The results identify the tonal complex-vowel segmented PSN (C_V_C_T) as the optimal model with a marginal R^2 of 0.162. The second highest ranking models belonged to two nontonal PSNs (C_V_C, C_G_VX) with marginal R^2 values of 0.154. An R^2 difference test showed that the C_V_C_T PSN was significantly higher than both competitors ($p < 0.001$).

Table 8: Model selection output for Experiment 2 according to marginal R^2 values for each model, and semipartial R^2 values for each fixed effect.

	CGVX	C_GVX	CG_VX	C_V_C	CG_V_X	C_G_VX	C_G_V_C	C_G_V_X
Model	0.133	0.135	0.114	0.154	0.108	0.154	0.124	0.121
Edit	0.004	0.005	0.005	0.004	0.004	0.005	0.005	0.004
Freq	0.002	0.018	0.028	0.021	0.026	0.038	0.033	0.032
HD	0.023	0.002	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
NF	0.001	0.005	< 0.001	0.004	< 0.001	0.008	0.001	< 0.001
k	0.043	0.004	0.010	0.003	0.002	0.017	0.003	0.001
CC	< 0.001	0.030	0.001	0.055	< 0.001	0.054	0.022	0.020
	+T	+T	+T	+T	+T	+T	+T	+T
Model	0.140	0.126	0.128	0.162	0.117	0.133	0.118	0.134
Edit	0.004	0.005	0.005	0.005	0.004	0.005	0.005	0.005
Freq	0.038	0.036	0.031	0.047	0.031	0.040	0.041	0.041
HD	0.008	< 0.001	0.006	0.005	0.005	0.005	0.005	0.007
NF	0.001	0.006	0.002	0.003	0.001	0.002	< 0.001	0.003
k	0.002	0.008	0.001	< 0.001	0.001	0.002	0.001	< 0.001
CC	0.021	0.003	0.022	0.060	0.012	0.025	0.009	0.011

Table 8 revealed that Freq according to tonal PSNs accounted for a greater portion of the variance than those of nontonal PSNs. NF played a limited role across all of the PSNs, while HD accounted for a portion of the variance in unsegmented and tonal PSNs (excluding C_GVX). Finally, despite *k* accounting for a portion of the variance for four of the PSNs (CGVX, CG_VX, C_G_VX, C_GVX_T), *CC* outranked *k* in semipartial R^2 for twelve PSNs.

The model estimates for the C_V_C_T PSN model, shown in Table 9, reveal that monosyllabic words greater in *CC* inhibited mental search and the production of phonological neighbors. Both high Freq and low Edit sped the search for neighbors. Tensor product smooths within a contour graph [106], as seen in Figure 4, were used to visualize a significant interaction between *CC*.C_V_C_T and Edit (adjusted R-sq. = 0.002; F = 11.85; $p < 0.001$). The graph reveals that contrary to the facilitative effect of low Edit, when the stimuli were low in *CC*, low Edit responses tended to be produced slower than high Edit responses.

6. Discussion

The model selection procedure used the lexical statistics tied to each of the sixteen PSNs to identify the likely structure used during mental search of phonological neighbors. Based on previous findings from the phonological association task of Wiener and Turnbull [70], we predicted a facilitative effect to high *k*. We also predicted the identification of an unsegmented PSN (CGVX, CGVX_T) based on the findings of production studies that hold that syllables are the first units for retrieval in Mandarin, i.e., "proximate units". Contrary to our predictions, model selection identified the tonal complex-vowel segmented PSN (C_V_C_T) without the expected facilitative effect of *k*. Interestingly, C_V_C_T was the same segmentation schema used to define phonological similarity in the Wiener and Turnbull study [70]. Meanwhile, the principle predictor within the C_V_C_T model, with a semipartial R^2 of 0.060, belonged to *CC* and was inhibitory in its effect on mental search.

The literature related to *CC* in the English mental lexicon entails inhibited retrieval and lower accuracy to high *CC* words. High *CC* has been tied to greater speech errors (Chan & Vitevitch 2010), lower accuracy in a perceptual identification task (Chan & Vitevitch, 2009), and the retention of newly learned nonwords (Goldstein & Vitevitch, 2014). Directly relevant to the current evidence is that high *CC* has also been shown to slow the retrieval of picture names (Chan & Vitevitch 2010), the judgment of lexical status of auditory words (Chan & Vitevitch, 2009; Goldstein & Vitevitch 2017) and visually presented orthographic words (Siew, 2018). The previous inhibitory *CC* findings are suggestive that our reaction times represent the selection of the target lexical item prior to production.

There are several indications as to why our results point to the C_V_C_T PSN. The first indication is the use of vowel information during the task. For example, glides, which are collapsed in this schema, were the least manipulated units in both experiments (Experiment 1: 2%; Experiment 2: 1%). A second indication is the length of our stimuli. Of the 198 stimuli in Experiment 2, nearly half consisted of three segments (SegLen 3 = 98). Through the disregard of the medial glides, which are obligatory in four-segment items, the 45 four-segment items were likely treated in the same manner as their three-segment counterparts. Given that 66% of all manipulations were of the substitution edit type, three-segment stimuli were primarily manipulated into three-segment responses. The final indication can be found in our participants' bias in producing tone neighbors (Experiment 1: 34%; Experiment 2: 46%). The influence of lexical tone was especially noted in the significant Freq effect across all tonal PSNs.

Of final concern is the significant effect of Edit. Participant-produced phonological neighbors that shared greater phonological similarity with the stimuli (i.e., lower

Table 9: Model estimates for Experiment 2.

Random effects		Variance	SD	Corr		
Item		0.001	0.035			
Subject		0.075	0.274			
Num_Chinese:Subject		0.003	0.055	0.590		
Residual		0.181	0.425			
Fixed effects	Estimate	SE	df	t value	p value	R^2
Intercept	1.536	0.066	32.15	23.22	< 0.001	
Edit	0.049	0.011	4713.00	4.67	< 0.001	0.005
Freq.C_V_C_T	-0.021	0.007	189.50	-3.08	0.002	0.047
HD.C_V_C_T	0.007	0.007	177.60	0.95	0.345	0.005
NE.C_V_C_T	0.006	0.007	183.30	0.78	0.438	0.003
k.C_V_C_T	0.001	0.008	191.30	0.14	0.891	< 0.001
CC.C_V_C_T	0.027	0.007	212.60	3.71	< 0.001	0.060

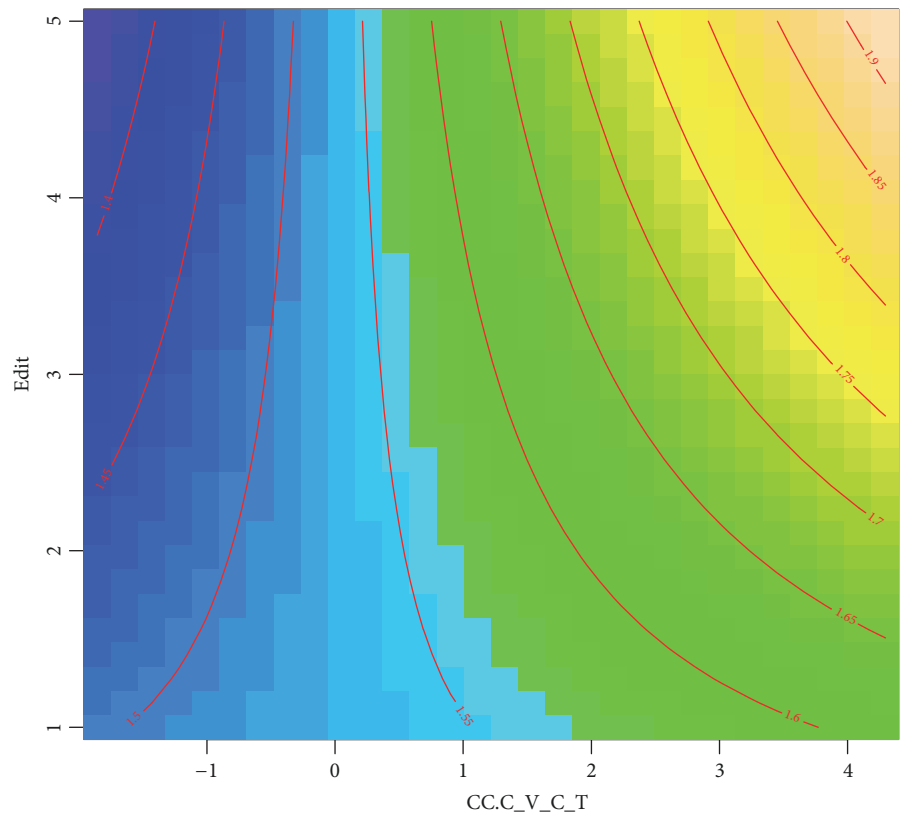


Figure 4: Tensor product smooth for the interaction between word-level clustering coefficient according to the tonal complex-vowel segmented PSN (*CC*.C_V_C_T) and the number of units (segments and/or tone) that differed between auditory stimuli and participant-produced phonological neighbors (Edit). Reaction times, as noted in the red contour lines, blend from cool (shorter latencies) to warm colors (longer latencies).

Edit) were produced faster than low similarity responses. These findings address a question posed by Vitevitch and colleagues [73] as to whether the edit distance between the stimuli and participant-produced phonological neighbors affects the time it takes to generate a phonological neighbor. In their study, they used neighbor generation as a means to investigate the types of neighbors that would occur to a given target if the target were incorrectly perceived. According to their hypotheses, our current result is suggestive that less time is needed to recover from the misperception of a spoken word when the misperceived item shares greater phonological similarity with its intended target.

7. Conclusion

In this study we constructed, measured, and then identified a possible schematic representation of phonological processing in the tonal language Mandarin Chinese. We began with

the identification of an optimal syllable inventory through 2 phonological association tasks. In Experiment 1, we used the edit distance between our participants' spoken responses and the annotation of the stimuli according to each syllable inventory to build and validate, (in Experiment 2), a novel syllable inventory that outperformed both prior inventories. On the premise of the nonuniqueness theory [84], the N&H inventory, built on phonological similarity, is the optimal choice to model Mandarin vocabulary in a phonological network in which relations between lexical items depend on phonological similarity.

The phonological association tasks aided in the identification of segmentation biases through spoken productions of phonological neighbors. Both tasks showed a strong tendency to use replacement as the method of manipulating units. The most commonly manipulated units were the items' lexical tones. In contrast, the most often ignored segments were medial glides.

The novel syllable inventory was used to build networks that we titled phonological segmentation neighborhoods (PSNs), in which schematic representations of segmentation determined phonological similarity. Each PSN was defined by it being built from one of sixteen phonological segmentation schemas. In using the same lexicon and number of nodes (30,000) within each PSN we were able to analyze the effects of segmentation and lexical tone on network statistics, both at the topological level and among monosyllables and disyllables. Segmented PSNs showed gradient differences according to the number of units within the syllable or whether or not they featured tone as a unit. For example, PSNs of less segmentation had greater $\overline{k}$ for both monosyllables and disyllables. $\overline{CC}$ did not show this pattern for monosyllables (except for the nontonal unsegmented PSN (CGVX_T), but did for disyllables, which is contrary to prior network findings [3, 5]. For nontonal segmented PSNs, the lack of tone also led to a greater $\overline{k}$ due to the mixing of syllable length.

The similarities between the sixteen PSNs and previous phonological networks were found in the presence of assortative mixing by degree and small world characteristics. The sixteen PSNs varied in Size, $\overline{k}$, $\overline{CC}$, $\overline{L}$, and whether or not they had power-law degree distributions. Discounting Size, three nontonal segmented PSNs, C_GVX, C_V_C, and CG_V_X, met all of the characteristics of the previously analyzed phonological networks. Contrary to our initial predictions, and those informed by the network analysis, our reaction time analysis revealed the tonal complex-vowel segmented PSN (C_V_C_T), with a significant inhibitory *CC* effect and a facilitative effect of low edit distance and high lexical frequency.

The current study began under the premise that the one-size fits all approach taken to phonological networks might not be sufficient. Yet, given the results of Experiment 2, do we have evidence to support this contention? The identification of the C_V_C_T PSN is likely the result of the stimuli we presented to the participants (majority 3 segments in length), and our participants navigation through the task demands, in that (1) the collapsing of vowel information in this PSN mirrors the lack of medial glide manipulations found in our participants' responses, (2) the primary method of substitution in order to produce a phonological neighbor meant that most productions were 3-segment neighbors of 3-segment stimuli, and (3) the presence of lexical tone in the featured PSN is likely the result of our participants' bias to use lexical tone as a guide through the mental lexicon.

The results are suggestive of complex adaptation wherein through manipulating the content and demand during a given task we identified the objects of those mental transformations. Thus far, network science methods have assisted both in the formation of the questions and how the results have been interpreted, despite the fact that the influence of topological features is still unclear. Future work will need to explore whether changes in the stimuli and task demands lead to the identification of different PSNs and whether those changes are meaningful representations of lexical processing.

Data Availability

The experiment data used to support the findings of this study have been deposited in github.com (https://github.com/karlneergaard/Constructing_the_Mandarin_phonological_network). The lexical database from which the network statistics were calculated has been deposited in github.com (https://github.com/karlneergaard/Database_of_word-level_statistics).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank the creators of the SUBTLEX-CH database, who provided the corpora used in the present study. We would like to thank Hongzhi Xu for his part in the creation of the lexical database and both Stephen Politzer-Ahles and Michael Tyler for their advice on the manuscript. The funding for this study was made available to the first author through a Hong Kong Polytechnic University (PolyU) International Postgraduate Scholarship and through the PolyU Faculty of Humanities International Collaboration project: 1-ZVKX Conversational Brains: A Multidisciplinary Approach.

References

[1] K. Y. Chan and M. S. Vitevitch, "The influence of the phonological neighborhood clustering coefficient on spoken word recognition," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 35, no. 6, pp. 1934–1949, 2009.

[2] M. S. Vitevitch, K. Y. Chan, and R. Goldstein, "Insights into failed lexical retrieval from network science," *Cognitive Psychology*, vol. 68, no. 1, pp. 1–32, 2014.

[3] K. Y. Chan and M. S. Vitevitch, "Network structure influences speech production," *Cognitive Science*, vol. 34, no. 4, pp. 685–697, 2010.

[4] R. Goldstein and M. S. Vitevitch, "The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition," *Frontiers in Psychology*, vol. 5, 2014.

[5] M. S. Vitevitch, K. Y. Chan, and S. Roodenrys, "Complex network structure influences processing in long-term and short-term memory," *Journal of Memory and Language*, vol. 67, no. 1, pp. 30–44, 2012.

[6] M. S. Vitevitch and N. Castro, "Using network science in the language sciences and clinic," *International Journal of Speech-Language Pathology*, vol. 17, no. 1, pp. 13–25, 2015.

[7] N. Castro, K. M. Pelczarski, and M. S. Vitevitch, "Using network science measures to predict the lexical decision performance of adults who stutter," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 7, pp. 1–8, 2017.

[8] C. S. Q. Siew, K. M. Pelczarski, J. S. Yaruss, and M. S. Vitevitch, "Using the OASES-A to illustrate how network analysis can be applied to understand the experience of stuttering," *Journal of Communication Disorders*, vol. 65, pp. 1–9, 2017.

[9] M. Stella and M. Brede, "Patterns in the English language: phonological networks, percolation and assembly models," *Journal of Statistical Mechanics: Theory and Experiment*, no. 5, 2015.

[10] M. Stella, N. M. Beckage, M. Brede, and M. De Domenico, "Multiplex model of mental lexicon reveals explosive learning in humans," *Scientific Reports*, vol. 8, no. 1, 2018.

[11] T. M. Gruenenfelder and D. B. Pisoni, "The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 3, pp. 596–609, 2009.

[12] R. Turnbull and S. Peperkamp, "What governs a language's lexicon? Determining the organizing principles of phonological neighbourhood networks," in *Proceedings of the 5th International Workshop on Complex Networks and their Applications (Complex Networks 2016)*, pp. 83–94, 2016.

[13] M. S. Vitevitch, "What can graph theory tell us about word learning and lexical retrieval?" *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 2, pp. 408–422, 2008.

[14] S. Arbesman, S. H. Strogatz, and M. S. Vitevitch, "Comparative analysis of networks of phonologically similar words in english and spanish," *Entropy*, vol. 12, no. 3, pp. 327–337, 2010.

[15] S. Arbesman, S. H. Strogatz, and M. S. Vitevitch, "The structure of phonological networks across multiple languages," *International Journal of Bifurcation and Chaos*, vol. 20, no. 3, pp. 679–685, 2010.

[16] W.-S. Lee, "A phonetic study of the "er-hua" rimes in Beijing Mandarin," in *Proceedings of the 9th European Conference on Speech Communication and Technology*, pp. 1093–1096, Portugal, September 2005.

[17] M.-F. Li, W.-C. Lin, T.-L. Chou, F.-L. Yang, and J.-T. Wu, "The role of orthographic neighborhood size effects in chinese word recognition," *Journal of Psycholinguistic Research*, vol. 44, no. 3, article no. 2, pp. 219–236, 2015.

[18] J.-T. Wu, F.-L. Yang, and W.-C. Jin, "Beyond phonology matters in character recognition," *Chinese Journal of Psychology*, vol. 55, no. 3, pp. 289–318, 2013.

[19] W. Wang, X. Li, N. Ning, and J. X. Zhang, "The nature of the homophone density effect: an ERP study with Chinese spoken monosyllable homophones," *Neuroscience Letters*, vol. 516, no. 1, pp. 67–71, 2012.

[20] W. Wen, "A study of Chinese homophones from the view of English homographs," *Modernizing our Language*, vol. 2, pp. 120–124, 1980.

[21] W.-F. Chen, P.-C. Chao, Y.-N. Chang, C.-H. Hsu, and C.-Y. Lee, "Effects of orthographic consistency and homophone density on Chinese spoken word recognition," *Brain and Language*, vol. 157-158, pp. 51–62, 2016.

[22] S. Xu, *Phonology of Standard Chinese*, Wenzi Gaige Chubans, Beijing, China, 1980.

[23] R. L. Cheng, "Mandarin phonological structure," *Journal of Linguistics*, vol. 2, no. 2, pp. 135–158, 1966.

[24] R. You, N. Qian, and Z. Gao, "On the phonemic system of standard chinese," *Zhongguo Yuwen*, vol. 5, no. 158, pp. 328–334, 1980.

[25] Z. M. Bao, "Fan-Qie languages and reduplication," *Linguistic Inquiry*, vol. 21, 1990.

[26] B. X. Ao, "The non-uniqueness condition and the segmentation of the Chinese syllable," *Working Papers in Linguistics*, vol. 42, pp. 1–25, 1992.

[27] S. Duanmu, *The Phonology of Standard Chinese*, Oxford University Press, Oxford, UK, 2nd edition, 2007.

[28] W. Lee and E. Zee, "Standard Chinese (Beijing)," *Journal of the International Phonetic Association*, vol. 33, no. 1, pp. 109–112, 2003.

[29] C. Shih and B. Ao, "Duration study for the bell laboratories mandarin test-to-speech system," in *Progress in Speech Synthesis*, J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, Eds., pp. 383–399, Springer-Verlag, New York, NY, USA, 1997.

[30] D. Wu, *Cross-Regional Word Duration Patterns in Mandarin*, University of Illinois at Urbana-Champaign, 2017.

[31] F. Wu and M. Kenstowicz, "Duration reflexes of syllable structure in Mandarin," *Lingua*, vol. 164, pp. 87–99, 2015.

[32] Y. R. Chao, *Eight Varieties of Secret Language Based on the Principle of Fan-qie*, 1931.

[33] M. Yip, "Reduplication and C-V skeleta in chinese secret language," *Linguistic Inquiry*, vol. 13, no. 4, pp. 637–660, 1982.

[34] P. G. O'Séaghdha and J.-Y. Chen, "Toward a language-general account of word production: the proximate units principle," in *Proceedings of the CogSci... Annual Conference of the Cognitive Science Society*, pp. 68–73, July-August, 2009.

[35] P. G. O'Seaghdha, J.-Y. Chen, and T.-M. Chen, "Proximate units in word production: phonological encoding begins with syllables in mandarin chinese but with segments in english," *Cognition*, vol. 115, no. 2, pp. 282–302, 2010.

[36] V. A. Fromkin, "The non-anomalous nature of anomalous utterances," *Language*, vol. 47, no. 1, p. 27, 1971.

[37] S. Shattuck-Hufnagel, "Speech errors as evidence for a serial-ordering mechanism in sentence production," in *Sentence processing: Psycholinguistic Studies Presented to Merrill Garrett*, W. E. Cooper and E. C. T. Walker, Eds., pp. 295–342, Erlbaum, Hillsdale, NJ, USA, 1979.

[38] J.-Y. Chen, "A small corpus of speech errors in mandarin chinese and their classification," *Word Chinese Language*, vol. 69, pp. 26–41, 1993.

[39] J.-Y. Chen, "The representation and processing of tone in mandarin chinese: evidence from slips of the tongue," *Applied Psycholinguistics*, vol. 20, no. 2, pp. 289–301, 1999.

[40] J.-Y. Chen, "Syllable errors from naturalistic slips of the tongue in Mandarin Chinese," *Psychologia: an International Journal of Psychology in the Orient*, 2000.

[41] A. S. Meyer, "The time course of phonological encoding in language production: phonological encoding inside a syllable," *Journal of Memory and Language*, vol. 30, no. 1, pp. 69–89, 1991.

[42] N. O. Schiller, "The effect of visually masked syllable primes on the naming latencies of words and pictures," *Journal of Memory and Language*, vol. 39, no. 3, pp. 484–507, 1998.

[43] N. O. Schiller, "Masked syllable priming of English nouns," *Brain and Language*, vol. 68, no. 1-2, pp. 300–305, 1999.

[44] N. O. Schiller, "Single word production in english: the role of subsyllabic units during phonological encoding," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 26, no. 2, pp. 512–528, 2000.

[45] J. D. Jescheniak and H. Schriefers, "Priming effects from phonologically related distractors in picture–word interference," *The Quarterly Journal of Experimental Psychology*, vol. 54, no. 2, pp. 371–382, 2001.

[46] A. S. Meyer and H. Schriefers, "Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 17, no. 6, pp. 1146–1160, 1991.

[47] J.-Y. Chen, T.-M. Chen, and G. S. Dell, "Word-form encoding in mandarin chinese as assessed by the implicit priming task," *Journal of Memory and Language*, vol. 46, no. 4, pp. 751–781, 2002.

[48] R. G. Verdonschot, M. Nakayama, Q. Zhang, K. Tamaoka, and N. O. Schiller, "The proximate phonological unit of chinese-english bilinguals: proficiency matters," *Plos One*, vol. 8, no. 4, 2013.

[49] W. You, Q. Zhang, and R. G. Verdonschot, "Masked syllable priming effects in word and picture naming in chinese," *Plos One*, vol. 7, no. 10, 2012.

[50] J.-Y. Chen, W.-C. Lin, and L. Ferrand, "Masked priming of the syllable in mandarin chinese speech production," *Chinese Journal of Psychology*, vol. 45, no. 1, pp. 107–120, 2003.

[51] T.-M. Chen and J.-Y. Chen, "The syllable as the proximate unit in mandarin chinese word production: an intrinsic or accidental property of the production system?" *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 154–162, 2013.

[52] A. W.-K. Wong and H.-C. Chen, "Processing segmental and prosodic information in Cantonese word production," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 34, no. 5, pp. 1172–1190, 2008.

[53] A. W.-K. Wong and H.-C. Chen, "What are effective phonological units in Cantonese spoken word planning?" *Psychonomic Bulletin & Review*, vol. 16, no. 5, pp. 888–892, 2009.

[54] A. W. Wong, J. Huang, H. Chen, and K. Paterson, "Phonological units in spoken word production: insights from cantonese," *Plos One*, vol. 7, no. 11, Article ID e48776, 2012.

[55] Y. Kureta, T. Fushimi, and I. F. Tatsumi, "The functional unit in phonological encoding: evidence for moraic representation in native japanese speakers," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 32, no. 5, pp. 1102–1119, 2006.

[56] Y. Kureta, T. Fushimi, N. Sakuma, and I. F. Tatsumi, "Orthographic influences on the word-onset phoneme preparation effect in native japanese speakers: evidence from the word-form preparation paradigm," *Japanese Psychological Research*, vol. 57, no. 1, pp. 50–60, 2015.

[57] K. Tamaoka and S. Makioka, "Japanese mental syllabary and effects of mora, syllable, bi-mora and word frequencies on japanese speech production," *Language and Speech*, vol. 52, no. 1, pp. 79–112, 2009.

[58] R. G. Verdonschot, S. Kiyama, K. Tamaoka, S. Kinoshita, W. La Heij, and N. O. Schiller, "The functional unit of japanese word naming: evidence from masked priming," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 37, no. 6, pp. 1458–1473, 2011.

[59] J. G. Malins and M. F. Joanisse, "Setting the tone: an ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese," *Neuropsychologia*, vol. 50, no. 8, pp. 2032–2043, 2012.

[60] J. G. Malins, D. Gao, R. Tao et al., "Developmental differences in the influence of phonological similarity on spoken word processing in Mandarin Chinese," *Brain and Language*, vol. 138, pp. 38–50, 2014.

[61] J. Zhao, J. Guo, F. Zhou, and H. Shu, "Time course of Chinese monosyllabic spoken word recognition: evidence from ERP analyses," *Neuropsychologia*, vol. 49, no. 7, pp. 1761–1770, 2011.

[62] J. G. Malins and M. F. Joanisse, "The roles of tonal and segmental information in Mandarin spoken word recognition: an eyetracking study," *Journal of Memory and Language*, vol. 62, no. 4, pp. 407–420, 2010.

[63] A. S. Desroches, R. L. Newman, and M. F. Joanisse, "Investigating the time course of spoken word recognition: electrophysiological evidence for the influences of phonological similarity," *Cognitive Neuroscience*, vol. 21, no. 10, pp. 1893–1906, 2009.

[64] S. Duanmu, "Chinese syllable structure," in *The Blackwell Companion to Phonology*, pp. 2151–2777, 2010.

[65] Y. H. Lin, *The Sounds of Chinese with Audio CD*, Cambridge University Press, 2007.

[66] X. Zhao and P. Li, "An online database of phonological representations for Mandarin Chinese," *Behavior Research Methods*, vol. 41, no. 2, pp. 575–583, 2009.

[67] A. Cutler, N. Sebastian-Galles, O. Soler-Vilageliu, and B. Van Ooijen, "Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons," *Memory & Cognition*, vol. 28, no. 5, pp. 746–755, 2000.

[68] A. E. Marks, D. R. Moates, Z. S. Bond, and V. Stockmal, "Word reconstruction and consonant features in english and spanish," *Linguistics*, vol. 40, no. 2, pp. 421–438, 2002.

[69] B. Van Ooijen, "Vowel mutability and lexical selection in english: evidence from a word reconstruction task," *Memory & Cognition*, vol. 24, no. 5, pp. 573–583, 1996.

[70] S. Wiener and R. Turnbull, "Constraints of tones, vowels and consonants on lexical selection in mandarin chinese," *Language and Speech*, vol. 59, no. 1, pp. 1–24, 2015.

[71] P. A. Luce and N. R. Large, "Phonotactics, density, and entropy in spoken word recognition," *Language and Cognitive Processes*, vol. 16, no. 5-6, pp. 565–581, 2001.

[72] M. Muneaux and J. C. Ziegler, "Locus of orthographic effects in spoken word recognition: novel insights from the neighbour generation task," *Language and Cognitive Processes*, vol. 19, no. 5, pp. 641–660, 2004.

[73] M. S. Vitevitch, R. Goldstein, and E. Johnson, "Path-length and the misperception of speech: insights from network science and psycholinguistics," in *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, A. Mehler, A. Lücking, S. Banisch et al., Eds., pp. 29–45, Springer, Berlin Heidelberg, Germany, 2016.

[74] M. Swadesh, "The phonemic principle," *Language (Baltim)*, vol. 10, no. 2, pp. 117–129, 1934.

[75] K. L. Pike, "Grammatical prerequisites to phonemic analysis," *Word*, vol. 3, pp. 155–172, 1947.

[76] W. J. M. Levelt, A. Roelofs, and A. S. Meyer, "A theory of lexical access in speech production," *Behavioral and Brain Sciences*, vol. 22, no. 1, pp. 1–38, 1999.

[77] G. S. Dell, "A spreading-activation theory of retrieval in sentence production," *Psychological Review*, vol. 93, no. 3, pp. 283–321, 1986.

[78] *I. Psychological Software Tools, "E-Prime 2.0."*, Psychological Software Tools, Inc., Pitsburgh, PA, USA, 2012.

[79] "汉典 zdic.net," http://www.zdic.net/.

[80] B. Li, *Research on Chinese Word Segmentation and Proposals for Improvement*, Roskilde University, 2011.

[81] X. Li, C. Zong, and K. Su, "A unified model for solving the OOV problem of chinese word segmentation," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 14, no. 3, pp. 12–29, 2015.

[82] Y. Zhang, J. Niehues, and A. Waibel, "Integrating encyclopedic knowledge into neural language models," in *Proceedings of the 13th International Workshop on Spoken Language Translation (IWSLT)*, 2016.

[83] J. Ma, C. Kit, and D. Gerdemann, "Semi-automatic annotation of chinese word structure and linguistics," in *Proceeding of the 2nd CIPS-SIGHAN Jt. Conference Chinese Lang. Process. (CIPS-SIGHAN 2012)*, vol. 0, pp. 9–17, 2012.

[84] Y.-R. Chao, "The non-uniqueness of phonemic solutions of phonetic systems," *Bulletin of the Institute of History and Philology Academia Sinica*, vol. 4, no. 4, pp. 363–398, 1934.

[85] Q. Cai and M. Brysbaert, "SUBTLEX-CH: chinese word and character frequencies based on film subtitles," *Plos One*, vol. 5, no. 6, Article ID e10729, 2010.

[86] M. Brysbaert and B. New, "Moving beyond kučera and francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english," *Behavior Research Methods*, vol. 41, no. 4, pp. 977–990, 2009.

[87] K. Neergaard, H. Xu, and C.-R. Huang, "Database of mandarin neighborhood statistics," in *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, pp. 2270–2277, May 2016.

[88] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *International Journal of Complex Systems*, vol. 1695, no. 5, pp. 1–9, 2006.

[89] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, Article ID 208701, pp. 1–5, 2002.

[90] M. E. J. Newman and J. Park, "Why social networks are different from other types of networks," *Physical Review E*, vol. 69, no. 3, pp. 1–9, 2003.

[91] J. M. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, no. 6798, p. 845, 2000.

[92] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," *Physical Review A*, vol. 38, no. 1, pp. 364–374, 1988.

[93] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[94] A. L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1, pp. 173–187, 1999.

[95] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.

[96] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *Society for Industrial and Applied Mathematics*, vol. 51, no. 4, pp. 661–703, 2009.

[97] C. S. Gillespie, "Fitting heavy tailed distributions: the poweRlaw package," *Journal of Statistical Software*, vol. 64, no. 2, pp. 1–16, 2015.

[98] P. Shoemark, S. Goldwater, J. Kirby, and R. Sarkar, "Towards robust cross-linguistic comparisons of phonological networks," in *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pp. 110–120, Berlin, Germany, August 2016.

[99] Y. H. Lin, *Autosegmental Treatment of Segmental Processes in Chinese Phonology*, University of Texas at Austin, 1989.

[100] I.-P. Wan, "A psycholinguistic study of postnuclear glides and coda nasals in mandarin," *Journal of Languages and Linguistics*, vol. 5, no. 2, pp. 158–176, 2006.

[101] M. E. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 98, no. 2, pp. 404–409, 2001.

[102] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[103] B. C. Jaeger, L. J. Edwards, K. Das, and P. K. Sen, "An R^2 statistic for fixed effects in the generalized linear mixed model," *Journal of Applied Statistics*, vol. 44, no. 6, pp. 1086–1105, 2017.

[104] B. Jaeger, *r2glmm: Computes R squared for Mixed (Multilevel) Models*, 2017.

[105] P. A. Jansen and S. Waiter, "Say when: an automated method for high-accuracy speech onset detection," *Behavior Research Methods*, vol. 40, no. 3, pp. 744–751, 2008.

[106] S. N. Wood, F. Scheipl, and J. J. Faraway, "Straightforward intermediate rank tensor product smoothing in mixed models," *Statistics and Computing*, vol. 23, no. 3, pp. 341–360, 2013.

Hindawi
Complexity
Volume 2019, Article ID 9424605, 27 pages
https://doi.org/10.1155/2019/9424605

Research Article

Expanding Network Analysis Tools in Psychological Networks: Minimal Spanning Trees, Participation Coefficients, and Motif Analysis Applied to a Network of 26 Psychological Attributes

Srebrenka Letina,[1,2] **Tessa F. Blanken**,[3] **Marie K. Deserno**,[4,5] **and Denny Borsboom**[4]

[1]*Department of Network and Data Science, Central European University, Hungary*
[2]*HAS Centre for Social Sciences "Lendület" Research Centre for Educational and Network Studies (RECENS), Hungary*
[3]*Department of Sleep and Cognition, Netherlands Institute for Neuroscience, Amsterdam, Netherlands*
[4]*Department of Psychology, University of Amsterdam, Amsterdam, Netherlands*
[5]*Dr. Leo Kannerhuis and REACH-AUT, Doorwerth, Netherlands*

Correspondence should be addressed to Srebrenka Letina; sreb.letina@gmail.com

Received 27 June 2018; Revised 16 October 2018; Accepted 29 November 2018; Published 21 February 2019

Guest Editor: Nicole Beckage

The analysis of psychological networks in previous research has been limited to the inspection of centrality measures and the quantification of specific global network features. The main idea of this paper is that a psychological network entails more potentially useful and interesting information that can be reaped by other methods widely used in network science. Specifically, we suggest methods that provide clearer picture about hierarchical arrangement of nodes in the network, address heterogeneity of nodes in the network, and look more closely at network's local structure. We explore the potential value of minimum spanning trees, participation coefficients, and motif analyses and demonstrate the relevant analyses using a network of 26 psychological attributes. Using these techniques, we investigate how the network of different psychological concepts is organized, which attribute is most central, and what the role of intelligence in the network is relative to other psychological variables. Applying the three methods, we arrive at several tentative conclusions. Trait Empathy is the most "central" attribute in the network. Intelligence, although peripheral, is weakly but equally related to different kinds of attributes present in the network. Analysis of triadic configurations additionally shows that the network is characterized by relatively strong open triads and an unusually frequent occurrence of negative triangles. We discuss these and other findings in the light of possible theoretical explanations, methodological limitations, and future research.

1. Introduction

In the last decade, network approaches have been increasingly used in psychological science for the investigation of psychological constructs and their interrelations in psychological science, as complementary or alternative to typically used and well-established methods (e.g., confirmatory factor analysis and structural equation modelling). This approach has introduced a different perspective on psychological constructs and has found its application in many subfields of psychology: intelligence [1], psychopathology [2], personality psychology [3], and social psychology [4]. One specific asset of the network approach is that it defines psychological constructs as constituents of a complex system of direct interactions enabling us to ask detailed questions about relationships of mutual influence among these constructs [5–8]. Specifically, Gaussian graphical models (GGM [9]) for continuous variables and Ising models for binary variables [10] have been used for network estimation with the aim to describe conditional independence relationships between variables, operationalized as partial correlations or conditional associations between variables [7, 11]. In this approach, a psychological network consists of nodes, psychological variables, and connections between nodes that represent the degree (and direction) of associations between each pair of variables, when the influence of every other variable in the network is controlled for.

After the construction of psychological networks, the quantitative analysis often proceeded with the computation

of a centrality analysis to answer which variable is most "dominant" or "important." Also, some global features have been of interest, such as network connectivity [12]. However, besides centrality measures and global measures of network structure, which focus on microscopic and macroscopic level of network, respectively, other analytical tools have been mostly ignored and rarely used in the study of psychological networks. This limited focus results in a limited set of questions that can be answered. We argue that, in order to answer research questions using psychological networks, researchers should go beyond the measures commonly used in psychology. The field of network science offers many alternative metrics that are worth considering when translating one's research question into quantifiable network properties. The main idea of this paper is to apply such techniques, which are already widely used in network science, to provide deeper understanding of psychological networks.

The structure of the paper is as follows. Firstly, we will describe some of the challenges in the analysis of psychological networks and link them with the three methods we propose in this paper, following with the general overview of the network that will be used for the demonstration of these methods. Next, we describe an illustrative dataset and apply the methods typically used in network analysis. Subsequently, we explain three methods that can be used to shed light on the network topology: minimum spanning trees (MSTs), the participation coefficient (PC), and motif analysis. For each method, we will explain specific procedures and modifications and conclude with results and discussion. Finally, in the general discussion, we summarize the benefits and possibilities of including the proposed methodologies in the field of network psychometrics and highlight interesting hypotheses that we arrived at using these analytical tools.

1.1. Identifying Challenges in the Analysis of Psychological Networks. In this paper, we propose three methods that not only provide novel insights into the network, but also circumvent some prominent methodological issues in the field of psychological networks: finding a way to operationalize the importance of all variables included in the network in a more general way, dealing with network of variables that are not of the same kind, and how to investigate the intermediate network level.

(1) Finding the hierarchical arrangement of nodes in the network: The main purpose of centrality analysis used in the analysis of psychological networks so far was to determine how entities in the network may be ordered regarding their connections with other variables (e.g., using the number and strength of connections) and regarding their overall position in the network, that is, to find out which entity is the most "dominant." The answers that arise from the application of different measures (typically strength, betweenness, and closeness) are likely to be different, as all of them capture different notions of what centrality means. However, the selection of the "right" measure is not the only challenge. Due to the small and dense nature of psychological networks, centrality measures may not meaningfully differentiate among specific nodes.

As a solution to those issues we suggest the use of the minimum spanning tree (MST), applied firstly on economics in the stocks analysis of time-series data [13]. The MST is a reduced subnetwork that connects all nodes based on the identification of the minimal set of edges needed. Besides providing a topological and hierarchically arranged skeleton of all nodes in the network, it additionally provides an insight into groupings of nodes based on their content similarity.

(2) The implicit assumption about the homogeneity of nodes in the network: Most commonly used centrality measures are based on a node's relation to every other node in the network. Thereby, these techniques implicitly assume that all nodes are a priori equally likely to be connected with any other node. This assumption is often untenable, as psychological networks may include one or more entities, or groups of entities, which differ in nature and/or measurement and therefore constitute a cluster (referred to as community or module, e.g. see [14]). In psychology, such a community may arise in part because of preexisting differences between the variables in, for example, nature of the variables (e.g., cognitive, behavioral, and emotional), kind of measurement (e.g., subjective vs. objective), or some methodological aspect of data collection.

In the estimated network, variables that are more similar regarding these preexisting differences (i.e., that belong to the same community in this sense) are more likely to be associated than variables belonging to different communities. Thus, these variables may show stronger associations among themselves and will *by construction* rank higher on common centrality measures like degree and strength. Note that this effect is especially pronounced when the size of different communities is not equal, as nodes belonging to the biggest community will by default have higher degree and strength. On the other hand, if some variables are different in some of the aforementioned ways from other variables included in the network, they may by default be expected to have less strong connections with other variables in the system. As a result, we might wrongly identify some node as central while, at the same time, a variable with a truly important role might be missed. This is important because psychological networks are increasingly starting to include psychological entities of different kinds. For example, recently some researchers [15] called for inclusion of other variables besides symptoms when analyzing psychopathological systems.

To circumvent the issue of nodes' heterogeneity, we propose the participation coefficient (PC [16]) to be used as a corrective in the procedure of estimating the most central node, because it addresses the uniformity of the edges a node has to different groups of nodes in the network.

(3) The network's mesolevel (or local structure): Visualization of small networks, such as psychological networks, provides immediate insights into the dyadic relationships between nodes, at the network as a whole, and even can provide some notion of the grouping of nodes. Similarly, measures typically used in the analysis of networks of this kind cover analyses at the microscopic network level. Macroscopic (global) properties of a network are easily computed, although their usefulness is less clear in psychological networks due to their small size and the impossibility to

claim that all relevant nodes are included in the network. The interpretation of commonly used centrality measures and global measures of network structure (e.g., average shortest path and clustering coefficient) as reflecting the importance of nodes in the system implicitly assume that the network contains all factors that are relevant to the system. However, one inherent characteristic of psychological networks is that it rarely models all factors that are relevant to the system [15]. In these cases, computing centrality measures based on indirect ties (betweenness and closeness) and global network measures may not capture all relevant information. While this is a problem when analyzing the entire system, much can be learned from shifting the focus to structural patterns on a more fine-grained level (i.e., mesoscopic level, "local" network structure). Methods for investigation of small configurations in network have been first developed in social network analysis [17] and have been redefined when applied to different types of (usually large) networks (e.g., neuronal networks, transcriptional networks, and the structure of the Internet) at the beginning of the century and have become known as "motif analysis." Motif analysis enables researchers to systematically investigate smaller configurations of nodes. It can help us determine, among other things, whether certain patterns, that is, subgraphs, represent interesting relations between constructs or methodological artefacts.

Moreover, this method addresses one of the basic questions in modeling networks: how global properties of networks can be understood from its local properties and how local topology is related to function [18]. For example, in psychological networks, different measures of intelligence are known to correlate positively; they show a positive manifold. In the language of network mesolevel analysis, this means that the system of different intelligence measures is characterized by smaller local structures that display positive relationships with each other. Van der Maas et al. [1] proposed a dynamical model of intelligence in which these patterns are interpreted as indicating that reciprocal causation or mutualism plays the most important process in that system. In other words, if a network expresses certain pattern of relationships in "high" degree, it may inform us about underlying process(es) driving the system that is represented as the network.

Each of the three methods, and especially the last two (the participation coefficient and motif analysis), give a clearer picture of *all* nodes in the network. It could be argued that they provide a *more "egalitarian" approach* to nodes that constitute a network, in a sense that they allow finding that noncentral (in terms of strength, betweenness, or closeness) nodes can be equally important for different parts of network or have an interesting role in a smaller part of network. That information can be easily overlooked when using only most basic network analytics. Given that psychological networks are usually relatively small, it is plausible that researchers will be interested to learn more about each node in the network, whether it is central or peripheral. Moreover, sometimes nodes that are peripheral can be of special interest and/or relevance (e.g., suicidal ideation in the network of depression symptoms and intelligence in the network of psychological traits).

1.2. Applying Three Methods in the Investigation of the Network of Different Psychological Attributes. Network analysis has been used mostly for looking more closely at one (or several related) psychological concepts, where nodes represent psychometric items that are part of a self-report measure (e.g., a questionnaire). In the current study, as an illustrative dataset for the proposed methods, we look at a network in which nodes are aggregated scores on self-report measures (also known as "parcels" of a questionnaire) that operationalize different psychological concepts (e.g., latent variables), most of which are not highly related, and among which direct causal relations may not be assumed. The variables in our network are supposed to measure relatively stable individual differences whose development "proceeds along mutually causal lines" [19, p.239]. Moreover, the conditional associations between those constructs are likely to be small, as most of them are assumed to be independent. To the best of our knowledge, this is the first research that looks at the network of different psychological attributes presented as aggregated items. We use network approaches to gain new insight into how different parts of that psychological system are connected, and which attributes have the most prominent role.

In the network of psychological constructs measured by self-reports we included cognitive ability (a proxy of g-factor [20]) measured with ability test (sometimes referred to in psychology as subjective and objective tests, respectively). The reason for including this substantially different variable in the network is twofold. First, we aim to demonstrate network methods that can provide more nuanced descriptions of all nodes, whatever their centrality in the network is. Including a variable, a node, which is known to be conceptually and methodologically different from others in the network, and at best only modestly associated with just some of nodes in the network, will set the stage for demonstrating added value of proposed methods. Second, we use the opportunity to address the old question of how cognitive ability and personality are related [21], to see how this question can be formulated and answered within the network approach.

Theoretically, intelligence is not expected to correlate with personality domain. For decades, researchers dealing with personality–intelligence connection have been using correlational studies to identify if significant relationship(s) exist(s). Yet, as Eysenck [22] in his review of the topic concludes, the research showed a striking lack of significant correlations, with few exceptions. For example, small associations have been found between intelligence and psychopathological profile [23], and introversion-extraversion related differences in style of intellectual performance (speed/accuracy ratio; [24]). Seeing that this approach failed to find any substantial relationship, Salovey and Mayer [25] suggest that question should be asked in a more complex way, for example, looking at the difference in the factorial structure of intelligence for groups with different personality profiles, and vice versa. Analytically, this suggestion is very much in line with network approach, because it looks at the whole set of variables at once, and is not as much focused on the size of specific effects. From a theoretical perspective, several attempts of an

integrative approach to both personality and intelligence with a wider theoretical framework for understanding their interrelations can be found in the literature, for example, social intelligence theory within cognitive theory of personality [26] and Motivational Systems Theory [27]. They are closely related to Smirnov's [28] view of intelligence as thinking, and personality as inherent component of all thought processes, while the link between the two is goals and problems in daily life.

2. Methods

2.1. Data and Measures. The dataset used in the current study has been collected within the context of the *myPersonality* project [29, 30]. In this project, participants self-administered one or more psychological questionnaires online, through a Facebook application (active from 2007 until 2012). Participation was voluntary and completely anonymous, and participants provided consent. In total, more than 20 different questionnaires were offered, and participants completed a self-chosen, variable number of questionnaires at a self-chosen place and time.

Of the available questionnaires, we selected 11 questionnaires, covering 31 psychological attributes, guided by three criteria: we wanted to include psychological concepts that (i) have a clear theoretical background and were measured with validated instruments with good psychometric properties; (ii) are considered to have high temporal reliability and stability; and (iii) had relatively high number (N>1000) of participants who also self-administered other questionnaires. To prevent including concepts that are too similar, we excluded concepts that correlated very highly to other concepts (around 0.60 in absolute correlations) and that had a clear theoretical overlap. This resulted in the inclusion of 26 psychological concepts. To facilitate interpretation, we reversed the scores of the negatively framed variables (Neuroticism, Depression, Militaristic values, and Violent occult interests) such that all variables can be interpreted as higher scores representing more favorable outcomes, except for Schwartz's values, where such rationale was not possible since having or not having high scores on certain value should be evaluation-free, meaning not positive or negative by default. The interpretation of the variables after recoding is listed in Table 1. More information on data processing, sample description, description of missing data, and descriptive statistics of 26 psychological variables is offered in the Supplementary Materials (SM, Sections 1-4).

We included 1,166,923 participants with a score on at least two of the psychological attributes (hereafter: variables). Of a subsample of participants, demographic information was available on gender (44.6%, of which 64.8% female and 35.2% male) and age (20.8%; M±SD = 26.1±6.7, range: 14-89 years). The sample consisted of participants from 220 different countries, and 35,7% of participants were from the US, UK, Canada, Australia, and India, respectively. A concise description of the included constructs and the instruments used is given in Table 1.

2.2. Network Estimation. We used partial correlations to estimate (For network estimation, visualization, and centrality analysis following R packages were used: BDgraph [42], qgraph [43], and networktools [44]. MST, PC, and motif analysis id done in NetworkX Python module [45]. Code used can be provided from the first author upon request.) the network. Partial correlation networks do not contain spurious correlations that are generated by common cause and chain structures within the network and can encode a basic data-generating network structure [46]. To estimate the network, we used a nonregularized method recently proposed by Williams and Rast (in press) [47] because, given our large sample size, relatively small number of variables, and our interest to detect weak ties, it is not advised to use regularization techniques like the LASSO that are often used ([47, 48], *in press*). More details about the process of determining the optimal estimation method for our data, and about the nonregularization method used, can be found in SM, Section 5.

To prevent the inclusion of spurious edges because of our overall large sample size, we artificially reduced the sample size by setting the N parameter in the estimation to N=4 131 (i.e., the median number of completed pairwise observations, for more details see SM, Section 3) instead of the total sample size of N=1 166 923. The estimated network is shown in Figure 1. The included edges were significant at alpha level of 0.001. Note that partial correlations are usually *smaller* than first-order correlations when interpreting the edge weights.

At first glance (Figure 1) at the network it can be seen that most of the nodes from the same group (questionnaire) cluster together in the network, except for Big Five traits that are more scattered across the network, especially Openness.

2.3. Robustness Analysis. To check robustness of our results, we tested it in two ways. First, we randomly split the sample in half 100 times and estimate a network on each half separately. Subsequently, we compare the two estimated networks on a metric of interest. If the network estimation is reliable, then the networks should be similar for both halves of the data, and, hence, the metrics should show high correspondence. This procedure is similar to that of Forbes et al. [49]. It should be noted, however, that, by using only half of the data to estimate a network, the statistical power drops considerably which will especially affect the estimation of small edges. Therefore, we conducted a second robustness analysis in which we randomly selected 100 sets containing 80% of the original sample and compared the network estimated on this subsample to the network estimated on the complete dataset. We computed the average correlation of the pairs of matrices estimated for the split halves (robustness analysis I) and between the whole sample and the random (80%) fractions (robustness analysis II). For the split halves, the average correlation was 0.82, indicating a high level of reliability. However, if we only evaluate the edges that are present in both estimates, on average, the reliability drops to 0.59 (similarity index). The average difference in the number of edges is 6.35, which is around 2% of all possible edges. For the random (80%) fractions, the similarity index increased to 0.85. The results are presented in more depth in SM, Section 6.

Table 1: Description of 26 psychological attributes included in the network.

Psychological attribute *(or Group of attributes (number of attributes in the group)), Questionnaire (author(s))*
Short description of measured attribute (number of items)
Values - based on Schwartz Theory of Basic Values (6), Schwartz Value Survey – SVS (Schwartz, 1992) [31]
Achievement - personal success through demonstrating competence according to social standards. (4 items)
Hedonism - pleasure or sensuous gratification for oneself. (3 items)
Power - social status and prestige, control or dominance over people and resources. (4 items)
Self-direction - independent thought and action—choosing, creating, exploring. (5 items)
Tradition - respect, commitment, and acceptance of the customs and ideas that one's culture or religion provides. (6 items)
Universalism - understanding, appreciation, tolerance, and protection for the welfare of all people and for nature. (8 items)
Big Five Traits (5), 20–100-item IPIP questionnaire (8 length versions), also included data on 336-item IPIP Personality Facets questionnaire (Goldberg et al., 2006). Both questionnaires are proxies for Costa and McCrae's NEO-PI-R facets (Five Factor Model) [32]
Emotional Stability (reversed Neuroticism) - the tendency not to experience negative emotions, such as anger, anxiety, or depression.
Extroversion - characterized by positive emotions, surgency, and the tendency to seek out stimulation and the company of others.
Openness to experience - a general appreciation for art, emotion, adventure, unusual ideas, imagination, curiosity, and variety of experience.
Agreeableness - tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others.
Conscientiousness - tendency to show self-discipline, act dutifully, and aim for achievement.
Interests (4), The Sensational Interests Questionnaire – SIQ (Egan et al., 1999) [33]
Low militaristic interests (reversed Militaristic interests) **–** an individual with low active interest in militaristic activities (e.g. guns and shooting). (10 items)
Low violent-occult interests (reversed Violent-occult interests**)–** an individual with low active interest in violent or occult activities (e.g. black magic). (7 items)
Intellectual interests – an individual's active interest in cerebral activities (e.g. philosophy). (7 items)
Interest in wholesome activities – an individual's active interest in active recreation (e.g. camping, hill walking). (5 items)
Body Consciousness (3), Body Consciousness Questionnaire –BCQ (Miller, Murphy, & Buss, 1981) [34]
Private body - awareness of internal sensations. (5 items)
Public body - awareness of observable aspects of body. (6 items)
Body competence – self-confidence in the body's performance. (4 items)
Integrity assessment (2), Rust's Sense of Fairness and Impression Management, Orpheus (Rust & Golombok, 1989), 36 items. [35]
Fair-mindedness (*or Sense-of-fairness*) – measures how balanced and impartial person is in her decision making.
Self-Disclosure – measures to what extent a person conducts her life transparently. Reversed values are used as a measure of Impression Management and Social desirability (Lie scale).
"Stand-alone" traits – six psychological attributes which are not part of a group of constructs, each is measured with separate questionnaire
Awareness of physical symptoms and sensations, *Pennebaker's Inventory of Limbic Languidness - PILL (Pennebaker, 1982)* [36]
Scale measures how often a person notices and reports a broad array of physical symptoms and sensations (e.g. chest pain, heart racing, dizziness). (54 items)
Self-monitoring, *Snyder's Self-Monitoring Scale, (Snyder, 1974)* [37]
Scale measures how much person monitors her self-presentations, expressive behavior, and nonverbal affective displays. (25 items)
Low Depression, *Center for Epidemiologic Studies Depression Scale (CES-D), NIMH, (Radloff, 1977)* [38]

Table 1: Continued.

Reversed Depression, measures lack of symptoms of depression in nine different groups as defined by the American Psychiatric Association Diagnostic and Statistical Manual, fifth edition. (20 items)
Empathy, *Empathy Quotient - EQ, (Baron-Cohen & Wheelwright, 2004)* [39] Scale measures self-reported ability to tune into how others are feeling, and to understand what they may be thinking. It measures both the affective and the cognitive components of empathy. (60 items)
Life satisfaction, *Satisfaction With Life Scale- SWLS (Diener, Emmons, Larsen, & Griffin, 1985)* [40] Scale measures general wellbeing and satisfaction with one's life. (5 items)
Intelligence, *MyIQ test,myPersonality's* 20-item proxy for Raven's Standard Progressive Matrices (Raven, 2008) [41], University of Cambridge's Psychometrics Centre (Chan & Kosinski) Ability test measures cognitive skills and clear-thinking ability, and pattern recognition abilities known to have the highest correlation with the general intelligence factor. (20 items)

Table 2: Description of ties in partial correlation network.

	Signed ties	*Absolute weights*	*Positive ties*	*Negative ties*
Mean	0,01	0,13	0,13	-0,13
SD	0,158	0,090	0,092	0,088
Min.	-0,39	0,05	0,05	-0,39
25%	-0,09	0,06	0,07	-0,16
Mdn	0,06	0,10	0,10	-0,10
75%	0,11	0,16	0,15	-0,06
Max.	0,53	0,53	0,53	-0,05
N. of ties	144	144	80	64

3. Illustrative Results: Network Description

3.1. Edge Weights in the Network. The current estimated network has 144 edges out of 325 possible edges, showing a good balance between sparsity and density (Figure 2). The distribution of the edges is summarized in Table 2; 64 edges (44%) are negative and 80 edges (56%) are positive. The number of negative edges is higher than usually observed psychological networks. Note that this is dependent on the network under consideration. If a network includes variables that all come from the same questionnaire (e.g., 10 depression items), then it would be expected that many (or all) edges are positive. In the current network, variables from various psychological questionnaires are included; they are not expected to correlate highly or/and positively by definition. Figure 2 also shows that, due to artificially decreasing statistical power and due to setting alpha to 0.001, edges around 0 are eliminated (< 0.05 in absolute value). For more details on the correlation network and estimated partial correlation network, and detailed analysis of ties, see Sections 7 and 8 in SM.

3.2. Centrality of Nodes. In addition to centrality measures that are typically used in psychological networks, we include more recently developed measures of node's expected influence ([50], for short explanation see Section 9 in SM).

Centrality measures can roughly be categorized into two groups, measures that look only at the local surroundings of a node (i.e., only the edges adjacent to the node) and measures that try to quantify the position of a node in the network by also taking into account nodes that are not directly adjacent to the node. Figure 3 shows centrality measures of the first category—considering only adjacent nodes. Figure 4 shows centrality measures of the second category—considering nodes beyond those directly adjacent to the node of interest. Comparing the different centrality measures, both within the same category or across categories, clearly shows that the measures diverge. Thus, different centrality measures indicate different nodes to be the most central. Although this follows logically from the way the different measures are computed, as each measure captures different aspects of centrality, it highlights the need to carefully consider the metrics used as it can greatly influence the answer to the question that is posed.

As Figure 3 shows, based on a node's direct ties, the most central node varies across measures. Based on strength, the value Tradition is the most central node, followed by Empathy, Extraversion, and another value, Universalism. Among the least central nodes are Agreeableness, Body Competence, and Awareness of Physical Symptoms.

Alternatively, when centrality measures consider more than the local environment of the node, a different arrangement of centrality emerges (Figure 4), with less agreement between different measures. Here, Empathy is the most central node, followed by Extraversion and Emotional Stability, while Tradition drops to the fourth place. The least central nodes are Self-Disclosure, Intelligence, and Awareness of Physical Symptoms.

Robustness analysis of all centrality measures used in this study is presented in Section 6 of SM.

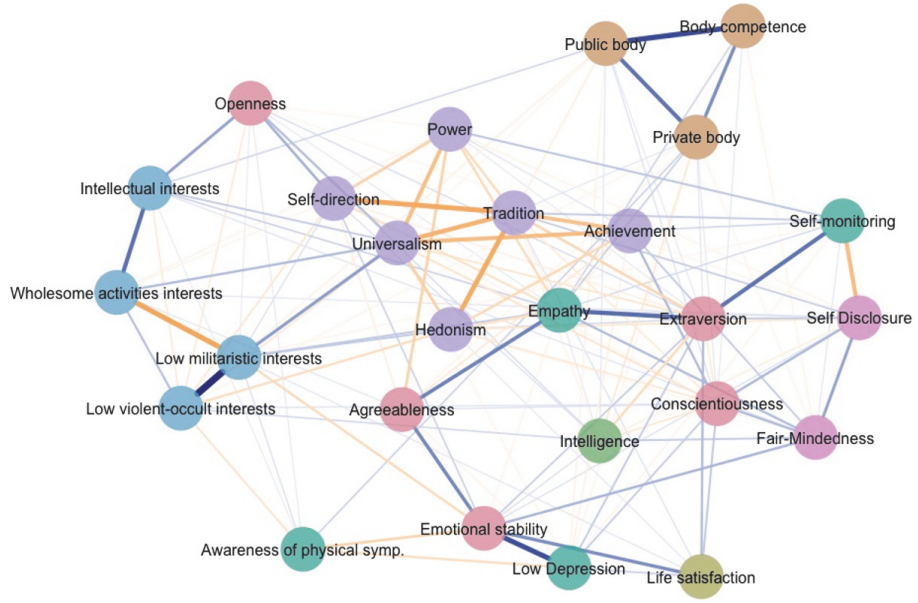


Figure 1: Nonregularized partial correlation network (set N= 4131, true N=1066921, layout spring, cut = 0). Blue ties signify positive relations and orange ties signify negative relations. The thickness of a tie is proportional to its absolute weight.

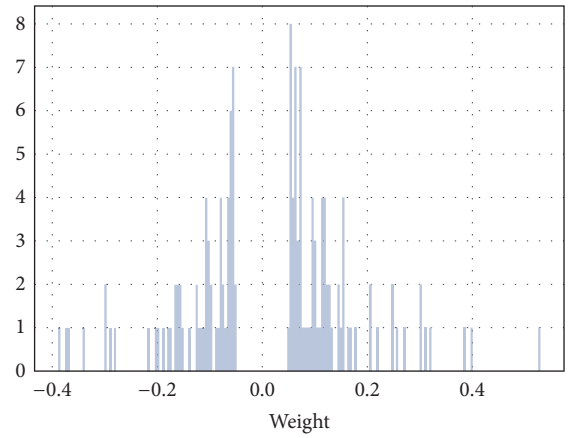


Figure 2: Distribution of weights in partial correlation network (L = 144).

4. Introducing Three Network Methods for the Analysis of Psychological Networks

4.1. The Minimum Spanning Tree. As demonstrated in previous section, different centrality measures capture different aspects of a node's position in the network, and the centrality of a node will differ depending on the centrality measure used. For that reason, we propose a way to look at the question about centrality differently, in a more general way. To be clear, we are not stating that centrality measures used so far in the research are inadequate, but we are merely trying to ensure a more general perspective to centrality. An alternative way to characterize relationship between all nodes in a network is by computing the minimum spanning tree (MST) [13]. The MST detects the hierarchical organization of the nodes and reduces the number of edges to those that carry the most information on the similarity of the nodes. Specifically, the MST is based on the distance between the nodes and selects the subset of edges (*number of nodes – 1*) without cycles, and with minimal total distance possible. This "skeleton" structure of the filtered network may be used if we want to get the answer to the general question which node is the most central, by not looking at the specific centrality aspects, but instead focusing on the network's most essential and local ties.

To compute the MST of our current network, first the distances among the nodes must be computed. An appropriate function for converting correlation to distances when negative correlations are present is as follows:

$$d(i,j) = \sqrt{2\left(1 - r_{ij}\right)} \quad (1)$$

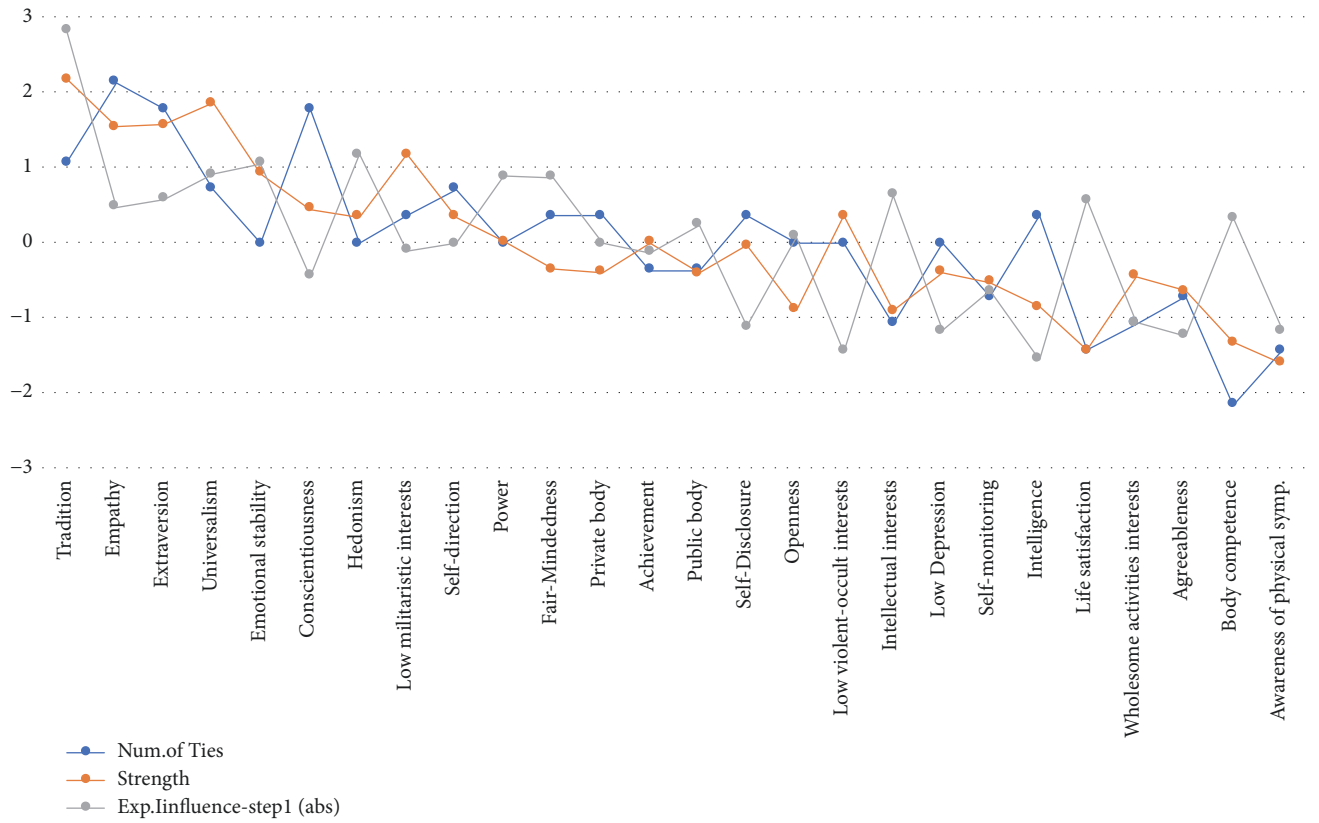


Figure 3: Centrality measures 1: based on node's direct ties (standardized values).

Equation (1) (Gower's distance measure [13, 51]) takes the direction of the correlation into account by assigning the largest distance to a perfect negative correlation, and the smallest distance to a perfect positive correlation. According to this equation the distances range from 0 to 2, where an intermediate distance of 1.4 is assigned to variables that are uncorrelated. The relationship between the (partial) correlation coefficient and the distance measure is shown in Figure 5.

Equation (1) is the preferred distance measure to distance inversely proportional to shared variance ($d(i, j) = 1 - pr_{ij}^2$). From the mathematical point of view, it is a more rigorous definition of distance and it gives monotonic transformation of coefficients. Most importantly, (1) gives more differentiated measure of distance than distance based on the shared variance, because in the latter the loss of information occurs since it translates partial correlations of the opposite sign and the same absolute values to the same distance. If negative ties are not present in the network, both measures will produce the same MSTs; otherwise the output will most likely differ (MST based on the shared variance is shown in SM, Section 11, Figure 15). Given the mentioned advantages and since almost half the ties in our network are negative, we have chosen to use it for MST construction. However, as it will be discussed in Section 5 and analysed in SM, Section 12, this measure is sensitive to reverse coding of variables included in the network.

Note that taking partial correlations instead of correlations when calculating distances means that, for each pair of nodes, it indicates how distant they are after the similarity based on covariance with other nodes in the network is excluded.

The MST of 26 psychological attributes is shown in Figure 6. The information about "centrality" of a node is very clear from the hierarchical structure, although centrality measures can provide more detailed picture (see SM, Section 10). The nodes with more direct edges and closer to the middle (centre) of the tree are most central.

Empathy is the most central node in the MST in the sense that it features the smallest distance to all other attributes. From Empathy, four branches emerge with only Sensational Interests and Body Consciousness being on the same branch as all other attributes from the same group. All branches are heterogeneous regarding the group of attributes they consist of, but they can be interpreted as having some commonalities in meaning. The branch with three Body Consciousness constructs along with Awareness of Physical Symptoms captures attributes related with body perception in general. The branch starting with Low Militaristic Interests can be interpreted as representing interests, values, and openness, which are related to what is often referred to as "lifestyle." The branch that starts with Extraversion relates to the attributes that describe one's agency and control in social world. Finally, the biggest and most heterogeneous

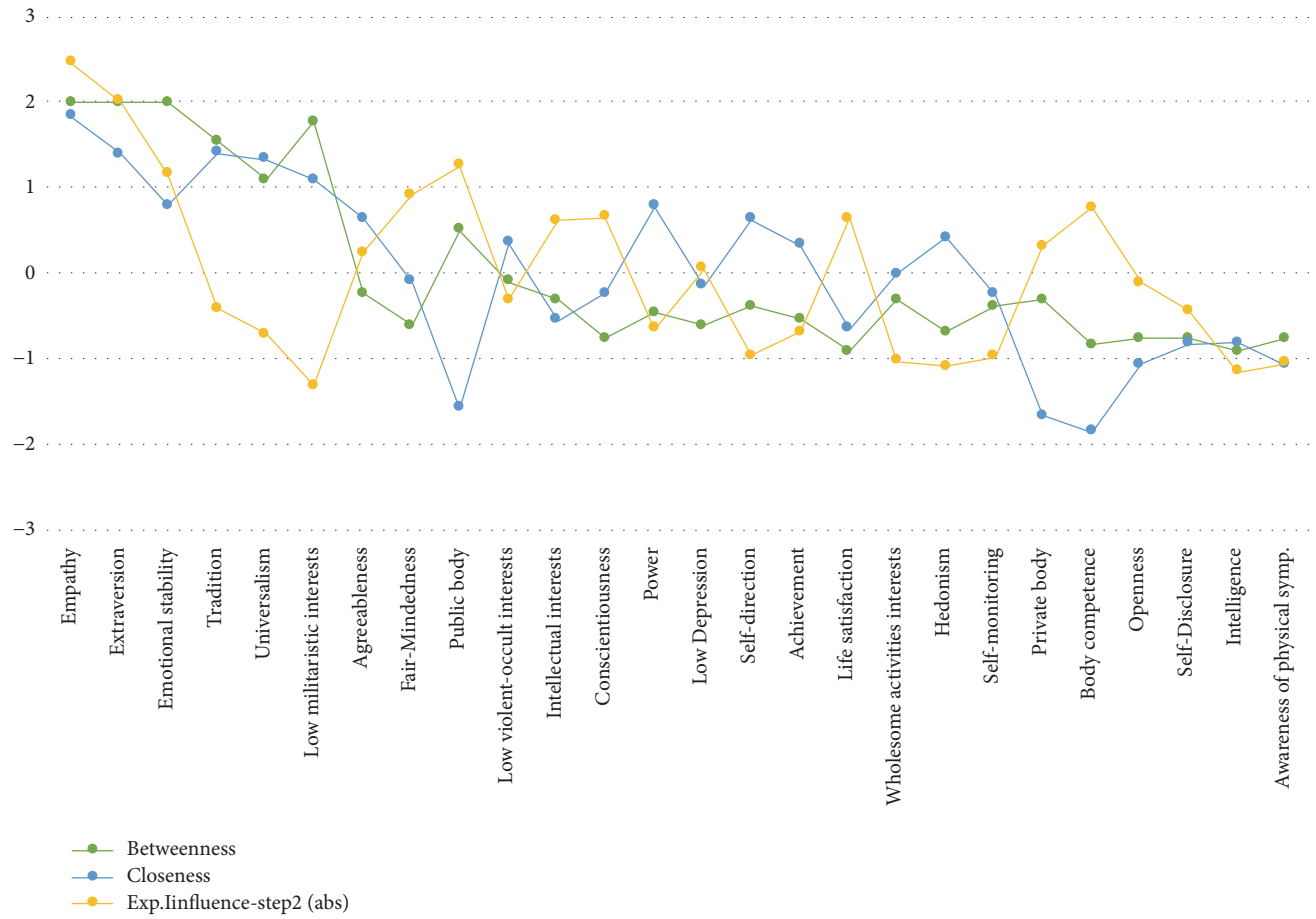


Figure 4: Centrality measures 2: based on links more than one distance away from the node (standardized values).

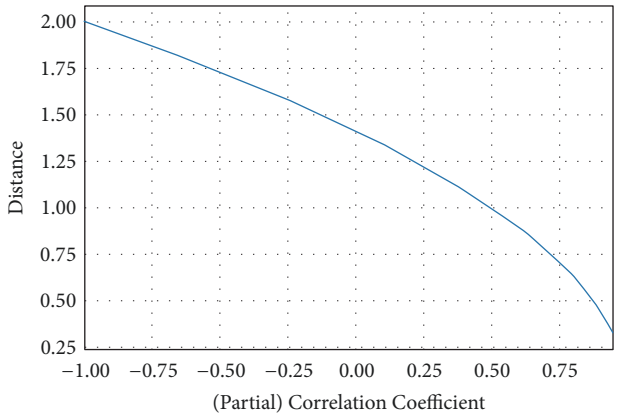


Figure 5: The relationship between partial or correlation coefficients and distance measure.

branch starting with Agreeableness is made of attributes that are highly socially esteemed and describe one's "relation" to others, oneself, and life in general. It is interesting to observe that Intelligence is placed on that branch and it branches out from Fair-Mindedness. This visual inspection shows another useful feature of MST; it gives indirect information on the hierarchical and overlapping, data-driven, clusters in the network. For example, in Figure 6, we can see two pairs of branches, or clusters, which *overlap* in Empathy. Alternatively, taking Empathy as the origin, there are four branches, or clusters, that overlap in that node.

According to the MST based on the distance defined in (1), two nodes are more distant in terms of steps (ties) between them in the filtered network (tree) if they are negatively associated than if they are not associated at all. That is why, for example, Tradition and Self-Direction (*pr*

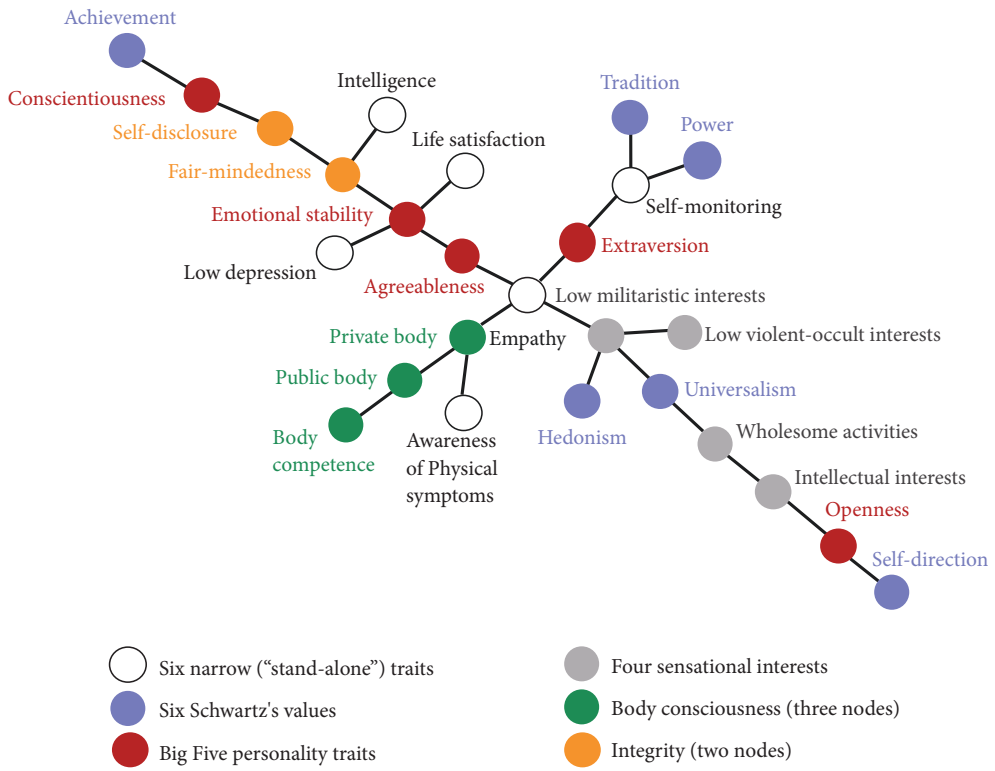


Figure 6: Minimum spanning tree (MST) based partial correlation.

= - 0,37) are placed on different branches and are more distant than Emotional Stability and Conscientiousness (*pr* = 0) that lie on the same branch. From the perspective of psychological networks, the MST preserves the specific content and meaning of the variable. More importantly, since its construction was affected by signs of weights, not only their absolute value, this filtered network can be a useful tool in testing whether two networks made of the same nodes really differ. Two networks estimated on two different samples will not usually be identical. However, if their MST is the same or very similar, this may indicate that their differences are not important. Similarity indexes of MSTs converge with the similarity indexes of whole networks. Nevertheless, reliability based on MST correlations seems to be lower than that based on network correlations in smaller samples (split halves), indicating that in fact the most informative ties are differently estimated (for details see Section 7 in SM).

4.2. The Participation Coefficient. In psychological networks, nodes (variables) may differ in their nature. Some may come from the same framework, while some may be stand-alone nodes. In network parlance, some nodes are part of a community and some nodes form a community of one or few. Note that these communities are not derived from data, but rather they are based on preexisting differences.

In the current dataset, for example, we had 26 psychological concepts, measured by 11 questionnaires. As such, there are groups of variables, varying in size, that belong to the same questionnaire and that are part of the same theoretical framework (e.g., three concepts on body consciousness) or measure the same kind of trait (e.g., measures of different "values"). Moreover, the psychological concepts that are part of the same questionnaire will likely be completed at the same time, while different questionnaires may have been taken days, months, or even years apart. Therefore, it is important to take these preexisting differences into account, if we want to explore which of the variables play an important role in the network.

One way to deal with these theoretically defined, preexisting communities is by employing measures that take this community structure into account and specifically evaluate connections a node has with nodes in different communities. One such method is the Participation Coefficient (PC), first introduced in the field of biological networks [16]. The PC takes the community structure into account, as it specifically quantifies how the edges a node has are distributed to different communities (similar in logic to Shannon entropy measure.). The important departure in our application of the PC is that it is not used on an empirical community structure, but rather on "communities," that is, groups of nodes and "stand-alone" nodes that were considered to exist in the network (a kind of "ground truth"). Framed as a hypothesis, the null hypothesis in the use of PC would state that preexisting groups of constructs (or data-driven communities) do not influence centrality scores of nodes. Showing that the rank order of nodes according to given centrality measure changes once the measure is corrected

with PC can be interpreted as supporting the rejection of the null hypothesis.

The calculation of the PC measure follows

$$PC_i = 1 - \sum_{m=1}^{G} \left(\frac{k_{i,m}}{k_i} \right)^2 \quad (2)$$

where PC_i signifies the PC score for a node i, while G, m, $k_{i,m}$, and k_i denote the network, each module in the network, number of ties of node i with nodes in that module, and number of all node's ties, respectively. The expression $k_{i,m}/k_i$ is simply the ratio of all node's ties that go to the specific module. In a version for weighted networks the number of links ($k_{i,m}/k_i$) in (2) is replaced with the sum of strengths which means that the expression $s_{i,m}/s_i$ signifies proportion of total strength of node i, invested in a single module:

$$PC_i = 1 - \sum_{m=1}^{G} \left(\frac{s_{i,m}}{s_i} \right)^2 \quad (3)$$

This difference means that if a node has the same number of links to every module, but they differ in strength, it will not achieve a maximum PC value. Here, strength is defined as the sum of absolute weights of all links involving node i, which means we disregard the sign of ties.

If a node has an equal number of edges to all the communities in the network (i.e., a uniform distribution of edges to all communities), the PC is closer to 1 (the highest possible value depends on the number of modules in the network; therefore average PCs of different networks can be compared only if PCs are normalized by theoretical maximum value, which is 0.50 for 2-module community, 0.80 for a network containing 5 communities, and, in our network containing 11 communities, the maximum PC value is 0.96). Alternatively, if a node has edges only to nodes within its own community, the PC is 0. It is important to note that the PC is not simply the number of links a node has to other communities in the network, but it rather quantifies the *equality of the distribution* of edges a node has to the other communities. In weighted networks, the PC is maximized if a node is connected equally to all the communities in the network: equal in both the number and the strength of edges to the other communities (i.e., a uniform distribution of edges and edge weights to all communities). More uniform distributions of nodes to all other communities correspond to higher PC values. For example, a node with one tie to each module will have the same PC as a node with two ties to each module. Similarly, a node with just one link to *each* module will have a higher PC than a node who has many links to some, but no links to other modules. A node with a high PC can influence all parts of the network *equally*, meaning that the node is equally important to every defined community. Such a node can be seen as a common denominator in terms of its potential influence on all communities in the network and can therefore help us understand the network as a whole. Note that PC considers only the node's direct ties, displaying the local perspective as MST. Moreover, that feature makes it very suitable for the analysis of a network where some elements of the network may not be included, and where therefore measures relying on the whole network (e.g., betweenness and closeness) may not be appropriate. However, since the PC solely quantifies the equality of the distribution of ties (or strength of those ties, in version for weighted networks) and disregards number (sum of strengths) of ties, we propose to use it in combination with a measure that considers both the number and the strength of the connections a node has and disregards the information about communities (preexisting or otherwise). One such measure is the Participation Ratio (PR [52]). Participation Ratio is defined [2] with the following formula:

$$C_D^{w\alpha}(i) = k_i \times \left(\frac{s_i}{k_i} \right)^{\alpha} = k_i^{(1-\alpha)} \times s_i^{\alpha} \quad (4)$$

where $C_D^{w\alpha}(i)$ is Participation Ratio of node (i), k_i is number of ties of node (i), s_i is the strength of the node, while α is a positive tuning parameter. If its value is set between 0 and 1, having a high number of ties (degree) increases $C_D^{w\alpha}(i)$, if $\alpha = 0$, it is equal to the node's strength, whereas, if α is set above 1, the number of ties decreases the value of $C_D^{w\alpha}(i)$, in such a way that a node with a greater concentration of its strength on only a few nodes and low degree has higher value than a node with the same strength but more ties. In our analysis, the α is set to 0.5, so that, for example, if a node A has a higher number of links and the same total strength as node B, the node A will have higher value of $C_D^{w\alpha}$. In this way both having high total strength and having more ties is favoured.

In short, PR is a single measure that quantifies both the number of edges a node has and the strength of these edges and weighs both equally (i.e., corresponding to an alpha of 0.5), and, as other measures defined so far in this paper, focuses only on node's direct links.

We transformed both measures to the same scale (range 0-1), visualized in Figure 7. Subsequently, for each node we computed the geometric mean of both measures. We opted for the geometric mean as it rewards consistency in scores on the two different measures. For each node, the PC, PR, and its geometric mean are shown in Figure 8.

Interestingly, as can be seen from Figures 7 and 8, the PC and PR can diverge for some nodes. For example, if we only focus on the number of edges and their strength, which is summarized in the PR, Tradition is highly central. However, Tradition has a relatively low PC, indicating that while it has relatively many and strong edges, these are not equally distributed throughout the network. Inspecting the estimated network in Figure 1, it can be seen that, indeed, the strongest edges of Tradition are mainly within its own community. Alternatively, Intelligence is not considered central based on the number and strength of its edges, but, taking the distribution of edges into account, we see that the connections of Intelligence are equally distributed to the other communities in the network (see Figure 16 in SM). This information would have been lost, if we had only focused on the number and strength of the edges (and other centrality measures related to these aspects).

In short, this example clearly illustrates that, when the objective is to find out which nodes play an important role in the network as connectors, it is important to consider whether there might be preexisting communities that should

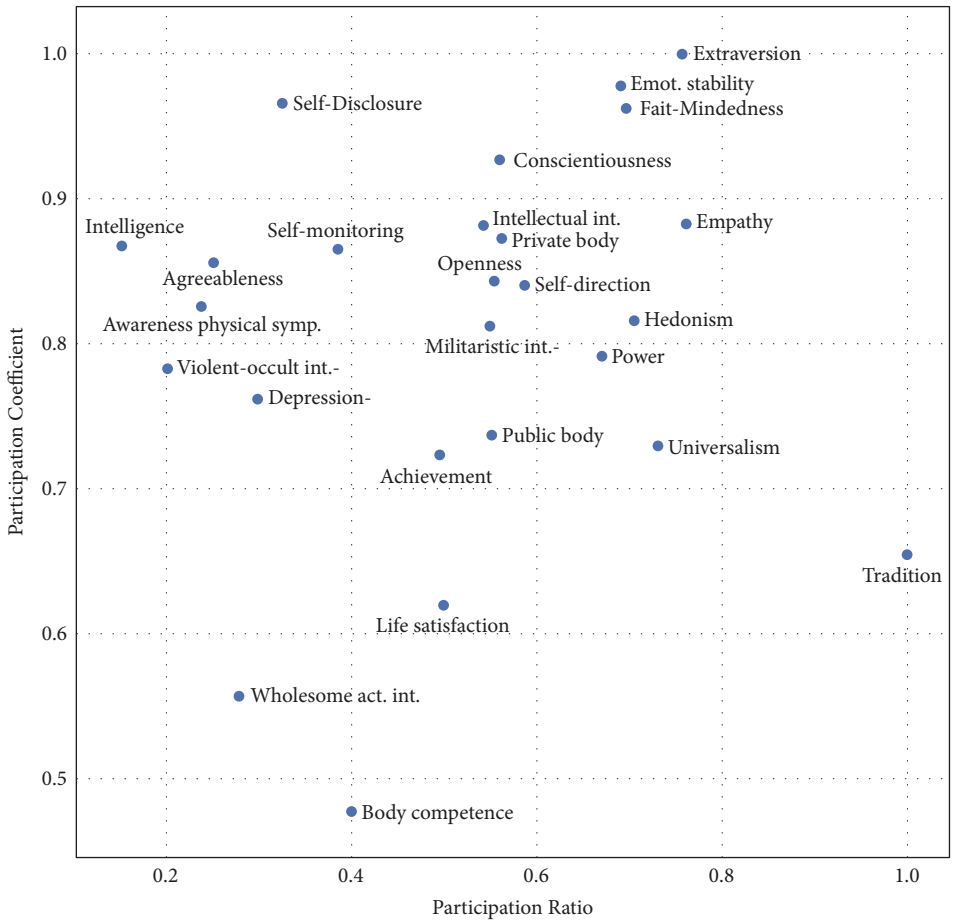


Figure 7: Scatterplot of standardized values of participation coefficient and participation ratio (min-max scale from 0 to 1) for all nodes in the network.

be taken into account. Not taking these preexisting communities into account might obscure the importance nodes belonging to small communities and "stand-alone" nodes that are not part of any community.

4.3. Analysis of Triadic Motifs. In this section, first we will explain the rationale behind the selection of motifs to be investigated, and the analysis of motif frequency, intensity, and coherence, followed by results and discussion, where the identification of specific motifs (and interpretation) is also included.

(i) Selection of motifs: Motifs usually represent subgraphs of three to five nodes for which different patterns of absent and present ties are examined. Many analyses of mesoscopic structures include or focus on triads, all possible configurations of three nodes. This is a sensible choice, because a triad is the smallest and the most basic network unit that defines the clustering of a network (transitivity) and can be characterized as the "simplest nontrivial motif" [53, p.2]. For undirected, unweighted, and unsigned networks, four types of triads exist: (1) triads without ties/edges (*empty triads*); (2) triads with one tie present, and two ties absent (*one edge triads*); (3) triads with one edge absent, and two edges present, referred to in the literature as *two-path*, *two-star*, or *open triads* (or *forbidden triads* in weighted networks when present edges are strong); and (4) triads with all edges present (*triangles*, *closed triads*) (Triads should not be confused with triplets. Triplets are like triads, but they are defined only by the presence of the edges and do not by the absence of edges. For example, both triangles and open triads are triplets of two edges.). Usually, the first two types of triads are not considered in the analysis, and some researchers define triads more strictly as systems of three nodes with at least two ties among them (e.g., [54]). The number of possible triads increases when the sign and weights of the edges are considered (e.g., [55]), as will be done in our analysis. Depending on the research question, some motif configurations may be of special interest and should be investigated, while others can be excluded from the analysis.

(ii) Analysis of motif occurrence, intensity, and coherence (including the identification of specific motifs): Once the motifs of interest are defined, the next step is to determine the frequency of each motif in the empirical network (each unique combination of three nodes is counted once). This yields a first insight into the network patterns at the mesolevel. The most frequent motif describes the most dominant pattern of connectivity in the given network among the motifs that are examined. However, the frequency alone yields

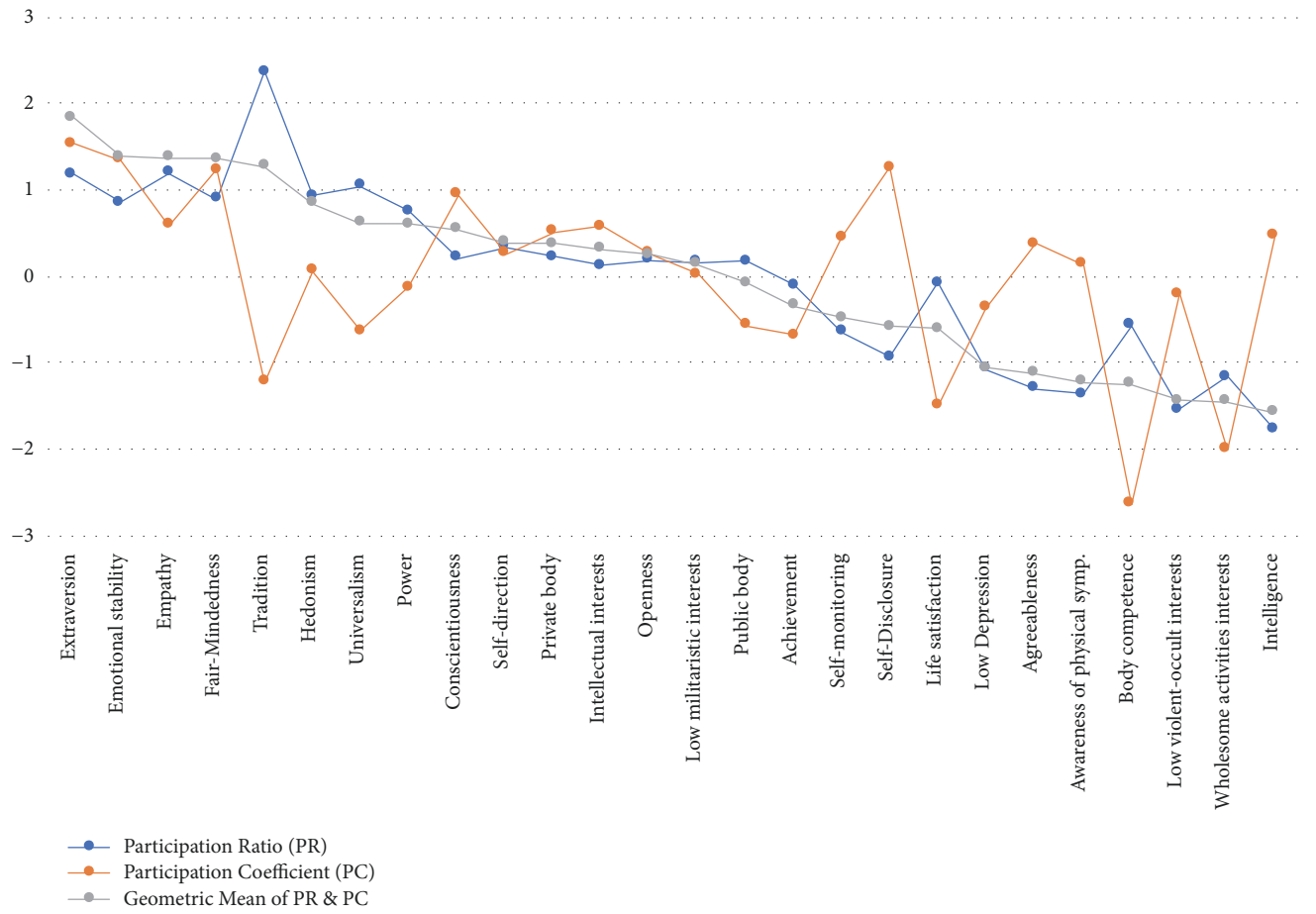


FIGURE 8: Centrality measures 3: participation ratio (the values of geometric means for Empathy and Extraversion are higher than both PR and PC. This is due to standardization of each measure. The plots with raw scores are shown in SM, Section 9, Figure 10.) ($\alpha = 0.5$), participation coefficient, and their geometric mean (standardized values).

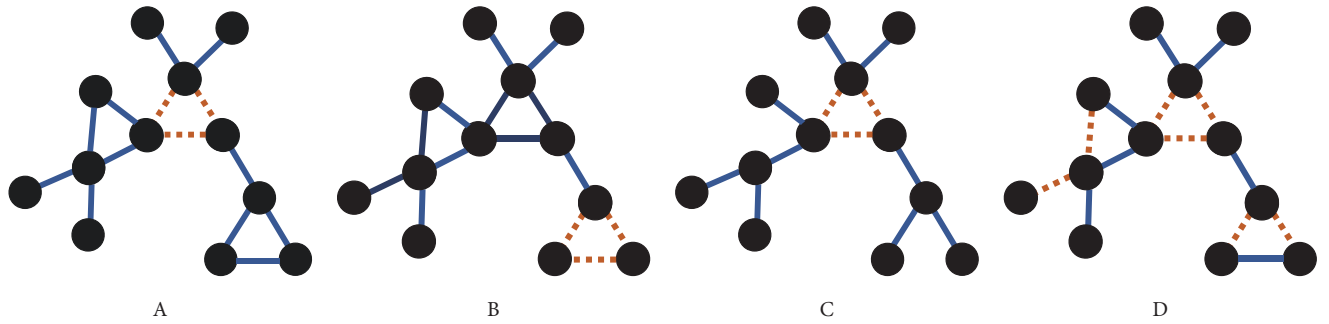


FIGURE 9: Four networks with 12 nodes and one negative triad (NNN). Networks A, B, and C have only three negative edges, while D has the same structure and density (number of edges) as A and B, but more negative edges.

limited information, because certain motifs might occur more frequently simply because of the network structure (in the context of describing the reference (null) model, the terms: network structure, topology, and degree sequence, are used interchangeably in this paper) and weight distribution. For example, imagine a hypothetical network of twelve nodes (variables) in which we observe predominantly positive edges, representing partial correlations between pairs of variables, except for three negative edges (described in Figure 9).

If we find one negative triad in a network, based on frequency alone, we could treat that finding as somewhat interesting but not especially informative about the network as a whole. However, when we consider what the chances are of observing three nodes connected with three negative edges in that system, that finding is of greater importance

for understanding the whole network as a system. Figure 9 describes extreme (and unlikely) examples of psychological networks which are used to illustrate why it is useful to additionally look at the chance of certain motifs occurring in the system. The weight distributions of networks A and B in Figure 9 are the same, while network C has a different structure compared to A, B, and D, because just one closed triad (triangle) is present. Since the structure is different, the weight distribution of C is also different. The chance of a NNN occurring in a network with the same structure and weight distribution is smallest in C, followed by A and B, where it is equal. The highest chance of observing such a triad is in D, because it has more negative edges and triads than other three networks. If networks are representing symptoms (behavioral, emotional, cognitive, or physical) of a disorder, three negatively associated symptoms in A, B, and *especially* in C are more important characteristic of the system than in network D. They are less likely to occur by chance in these three networks, and therefore more likely to describe a process which is important for understanding the network. For example, a triad NNN in A could be interpreted as a process of negative feedback which is central for the network (it "drives" the network). In B, NNN is equally important but it describes occurrence of a negative "loop" in a peripheral part of the network, among symptoms that are less central. In C, NNN is even more essential for understanding the network than in A, as it could be described as the sole driving force of the network, each of the negatively connected nodes in the triad relates to a different set of nodes. Note that motif analysis per se does not differentiate between A and B as the centrality of configurations is not accounted for. Finally, NNN in D is a central configuration which shows an interesting pattern of association between three symptoms, worth of attention in the interpretation of the network. However, it is not as important for describing the process underlying the network formation since other negative associations between nodes and within triads are present. The same reasoning applies if nodes are representing other nonpathological tendencies, like personality traits, values, etc. In these networks the difference will be in the average weights of edges, which is likely to be smaller than in case of networks featuring psychopathological symptoms or other more correlated variables.

Therefore, for each motif, we establish whether it occurs more or less frequently than would be expected by a null model. In weighted networks, the appropriate null model is a random network (to be precise, it is not a random graph model, but a configuration model (for more details see [56])) with fixed topology (degree sequence) and randomized weights from the same distribution of weights as observed in the empirical network (for more details on general null models see [57]). The quantification of occurrence of a specific configuration in a network is usually done by comparing it with the occurrence of the same motif in a reference model (for introduction see [18]). Distribution of motif frequencies is obtained by generating a sample of random networks. The empirical frequency of a motif is compared against that distribution and if it appears significantly (this significance should not be confused with significance of ties in the motif) more (less) often than it would be expected by reference model it signifies the motif is indeed "a motif." (Sometimes the term "motif" is used only for these configurations for which this step of analysis shows that they are significantly over- or underrepresented. In this article, we do not make such distinction, as we refer to every investigated configuration as a motif, and after the analysis is done, we describe it as significant or not.) Itdescribes an important characteristic of the investigated network. Motifs that occur more frequently describe a common configuration of nodes and therefore provide information about the network connectivity. Moreover, these motifs could have some important functional roles in the system. For example, closed triads are usually overrepresented in social networks, because they represent a process of social (triadic) closure, while in a network of intelligence measures they may indicate the process of mutualism [1].

However, in weighted networks the analysis of motif frequency omits the information about the weights (unless it is in some ways included in the definition of the motif). For example, if two motifs have the same occurrence in a network (let us assume for the sake of the argument that both have equal distribution of frequencies based on appropriate random models), but the first is (on average) made of stronger ties than the second, we cannot treat them as equally describing the local structure of the network, that is, to be equally likely to describe some important process in the network. Although they are equally present in the network, the first is expressed more strongly and is therefore more likely to describe some important process.

To address this issue, Onnela et al. [53] introduced the Intensity measure (the geometric mean of all the weights (in the case of absent ties in the motif, these are treated as zero weights) in a motif ((5), where l_g stands for number of ties in the motif)), which looks at the motifs not as discrete objects who are either present or not (expressed or not expressed) in the network, but rather as objects existing on a continuum, where zero or low Intensity values imply that motif is present in low degree. As such, the Intensity I can be used to identify high and low Intensity motifs in the system:

$$I_{(g)} = \left(\prod_{(i,j) \in l_g} w_{ij} \right)^{1/|l_g|} \tag{5}$$

In addition to Intensity (I), a Coherence ($Q_{(g)}$) ratio can be computed that quantifies how internally coherent the weights in motifs are by computing the ratio between the geometric and the arithmetic mean. It ranges from 0 to 1, with higher scores indicating less difference between the weights (in absolute terms). As was the case with the analysis of occurrence of motifs, the significance of both Intensity and Coherence is estimated in comparison with the distribution of their values for a given motif in reference model.

A motif that is underrepresented in the network, in terms of occurrence or intensity, describes a pattern of relationships which, for some reason, is unlikely to happen in a network. In other words, when we exclude the hypothesis that a given occurrence or intensity of a certain configuration does not come from a reference system, it points out that there may

be an additional origin for the effect, possibly the function of the system [18]. In case of psychological networks, the occurrence and significance of a motif which is not easily interpretable may also happen as an artefact (e.g., due to the sample on which the network is estimated, problems in the network estimation procedure, or measurement error). For that reason, a motif analysis can be useful in the analysis of psychological networks, forasmuch as it can help quantify and identify presence of unexpected configurations in the network as well.

In the next section, the motif analysis on illustrative data is described in detail and results are presented and discussed.

4.3.1. Selection of Motifs and Analysis of Motif Occurrence. When the sign of an edge is considered, seven configurations of triads are possible (disregarding empty triads and triads with only one edge, see Figure 9). Four of them fall under "closed" triads or triangles: triads with either only positive (positive triad, PPP) or only negative weights (negative triad, NNN) and triads consisting of two positive and one negative weight (PPN) or two negative and one positive weight (NNP). NNN and PPN are also known as imbalanced triads (NNN is also sometimes considered as imbalanced triad in social networks, but some debate exists over whether it is truly imbalanced or not. Not to confuse with too many similarly named triad, we will use the term "imbalanced" in this article only when referring to triad with one positive and two negative ties and to triads that do not satisfy the triangle inequality principle (the latter is explained in the following text)), in social balance theory [58, 59] because they signify configurations of affective ties between persons which is not likely to appear in social networks (or if it appears it is not likely to persist; that is, it is likely to change). The remaining three triads are open triads (2paths) consisting of two ties: with only positive weights (2path pos., P0P where "0" stands for the absent weight), only negative weights (2path neg., N0N), or with one positive and one negative weight (2path mixed, P0N or N0P).

Networks, especially social networks, tend to show transitivity; if person A is connected with (friend of) person B, who is connected with (friend of) person C, A and C are likely to be connected (friends). Although, in recent years, we have witnessed a surge of research on psychological networks, we still do not know enough about their general properties. Correlations, and especially partial correlations, do not have to be transitive, but it is often the case that if a trait A positively correlates with trait B, which is also correlated positively with trait C, then we expect traits A and C to correlate positively as well. If that is the case, P0P motifs should appear less often than expected by the reference model. Likewise, according to the social balance theory, closed triads with one or three negative edges (i.e., PPN and NNN) are less likely to occur in social networks [58–60]. We hypothesize that, in psychological networks too, NNN and PPN triads represent configurations which are not expected to occur frequently because of two reasons. First, it is challenging to explain how three psychological attributes feature negatively partial correlations. One possibility is that a process of negative feedback among attributes exists. A second possibility is that the three nodes positively contribute to a common effect, which has been implicitly or explicitly conditioned on. A third possibility is that the variables are measured with error, and the partial correlation picks up negative correlations between the error terms.

On the other hand, positive associations between A and B, and B and C, render a possible negative association between A and C difficult to interpret (PPN triad). The importance of detecting such configurations in psychological networks lies in the fact that they either describe unusual finding(s) or may point to the existence of methodological artefacts. In both cases, we benefit from knowing about the presence of such configurations. It should be noted that, while it is more straightforward to predict that such configurations could be less frequent in a correlation network, in the case of partial correlation network they could be more likely to occur. To the best of our knowledge no analysis of this kind has been performed on a network representing (partial) correlations. The summary of hypotheses is shown in Table 4, in Section 5.

Among the motifs (Figure 10, third row), the only significant motif is the negative triad (percentile 99.7). In other words, the negative triad appears more frequently than would be expected by chance, given the same degree sequence and weight distribution. Path2 with positive ties (P0P), indicating high presence of nodes which are bridges, is overrepresented, and the imbalanced triad (PPN) is underrepresented, but neither reaches the level of significance.

To identify only the strongest motifs, we looked at signed motifs with an added threshold (see Figure 11). To end up with a similar number of examples for each motif, we selected a threshold of 0.15 (around 75 percentile of edge weights, see Table 1) for closed triads and a threshold of 0.20 for 2path motifs. Among the motifs that meet this threshold, one specific motif may be of relevance for psychological networks. This is the last motif in Figure 11, which we called imbalanced triplets II T., based on the work of Toivonen et al. [61] (hence the T. in the name, for definition see Figure 11). Toivonen and colleagues investigated a correlation network of emotion concepts and argued that this motif describes patterns that cannot be depicted in any dimensional space without being distorted. This "imbalanced triplet" describes a pattern which is contraintuitive, although not necessary unreal, and it is similar in logic to NNP triad. If A, B, and C represent three psychological dimensions (e.g., emotions and traits), and positive correlations between A and B, and B and C exist, depending on the strength of r_{AB} and r_{BC}, A and C ought to correlate at least as the half of either of the two (r_{AB} or r_{BC}) which is the weaker correlation. Otherwise the ABC triad does not satisfy the triangle inequality principle; that is, it cannot be described by dimensional techniques (in Euclidean space), while a network representation can be used for detecting their presence.

As mentioned for the NNN and PPN motifs, while we can expect low occurrence of imbalanced triplets II T. in a correlation network, in a partial correlation network this is quite different. An imbalanced triad in a partial correlation network implies that the partial correlation between A and B is small given C, which means that A and B approach

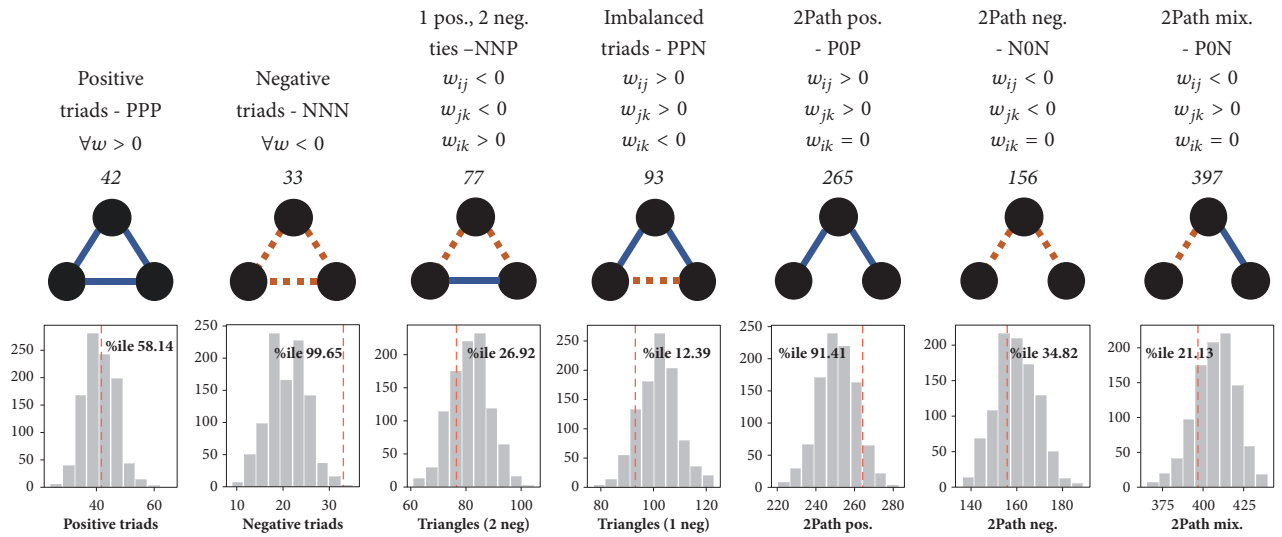


FIGURE 10: Signed motifs, name used in this study, the definition, schematic figure, and the figure showing distribution of motif frequencies in 1000 random networks with the same degree sequence and weight distribution and percentile value of the frequency of empirical network in that distribution.

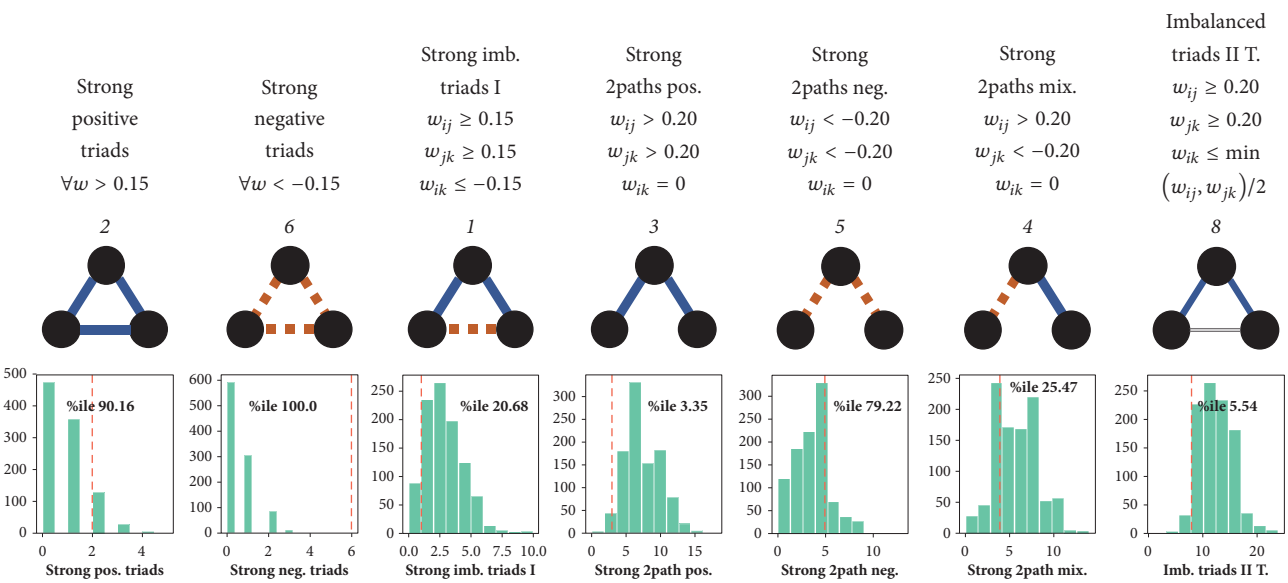


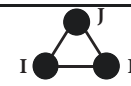
FIGURE 11: Weighted and signed motifs, name used in this study, the definition, schematic figure, and the figure showing distribution of motif frequencies in 1000 random networks with the same degree sequence and weight distribution and percentile value of the frequency of empirical network in that distribution.

conditional independence given C. This in turn is consistent with a chain (A->B->C or A<-B<-C) or a fork (A<-B->C). Both of these may yield indirect, but important clues to the causal structure within the triad. Those triads are good candidates for more focused analytical approaches that allow for causal inference (e.g., mediation or path analysis). Thus, regardless of frequency, the imbalanced triplets II T. represent a configuration that describes possibly interesting phenomena which would go unnoticed with dimensional methods [61].

Results show that, even when we "focus" just on motifs of relatively strong ties (Figure 11 (third row), all of them identified in Table 3) again only the NNN triad occurs significantly and more than expected by chance. The cardinality (a term used in network analysis to address the significance of a motif) of the motifs in this network is thus not dependent on the strength of the weights. However, the strong imbalanced triads, 2paths with positive weights, and imbalanced triplets II T. have the tendency to be underrepresented. This pattern is expected in social networks, where imbalanced triads and "forbidden triads" (2paths) are generally less expressed, and this network shows similar tendencies.

All motifs defined in Figure 11 are identified and described in more detail in Table 3.

Table 3: Weighted and signed motifs identified.

Motif	 i, j, k; **(pr_{ij}, pr_{jk}, pr_{ik})**, [r_{ij}, r_{jk}, r_{ik}]
	Private body, Public body, Body competence; **(.30, .40, .25)**, *[.52, .58, .49]*
	Conscientiousness, Fair-Mindedness, Self-Disclosure; **(.16, .21, .16)**, *[.34, .37, .37]*
	Tradition, Universalism, Power; **(-.34, -.29, -.16)**, *[-.20, -.46, -.11]*
	Tradition, Universalism, Hedonism; **(-.34, -.21, -.36)**, *[-.20, -.12, -.38]*
	Tradition, Universalism, Achievement; **(-.34, -.30, -.28)**, *[-.20, -.34, -.24]*
	Tradition, Self-direction, Power; **(-.37, -.20, -.16)**, *[-.47, -.12, -.11]*
	Tradition, Hedonism, Achievement; **(-.36, -.17, -.28)**, *[-.38, .01, -.24]*
	Universalism, Hedonism, Achievement; **(-.21, -.17, -.30)**, *[-.38, -.01, -.24]*
	Militaristic int.-, Universalism, Wholesome act. int.; **(.22, .16, -.39)**, *[.20, .19, -.37]*
	Agreeableness, Empathy, Extraversion; **(.27, .32, .0)**, *[.45, .39, .16]*
	Life satisfaction, Emotional stability, Agreeableness; **(.25, .26, .0)**, *[.48, .35, .25]*
	Wholesome act. int., Intellectual int., Openness; **(31, .21, .0)**, *[.41, .44, .17]*
	Self-direction, Tradition, Universalism; **(-.37, -.34, .0)**, *[-.47, -.20, .21]*
	Hedonism, Tradition, Self-direction; **(-.36, -.37, .0)**, *[-.38, -.47, .16]*
	Achievement, Tradition, Self-direction; **(-.28, -.37, .0)**, *[-.24, -47, .09]*
	Self-direction, Power, Universalism; **(-.20, -.29, .0)**, *[-.12, -.46, .21]*
	Hedonism, Universalism, Power; **(-.21, -.29, .0)**, *[-.12, -.46, .21]*
	Militaristic int.-, Universalism, Tradition; **(.22, -.34, .0)**, *[.20, -.20, -.10]*
	Intellectual int., Wholesome act. int., Militaristic int.-; **(.31, -.39, .0)**, *[.41, -.37, -.11]*
	Militaristic int.-, Universalism, Power; **(.22, -.29, .0)**, *[.20, -.46, -.10]*
	Militaristic int.-, Universalism, Achievement; **(.22, -.30, .0)**, *[.20, -.34, -.10]*
	Self-monitoring, Extraversion, Empathy; **(.30, .32, .09)**, *[.31, .39, .12]*
	Depression-, Emot. stability, Agreeableness; **(.38, .26, -.06)**, *[.55, .35, .17]*
	Emot. stability, Agreeableness, Empathy; **(.26, .27, -.06)**, *[.35, .45, .15]*
	Violent-occult int.-, Militaristic int.-, Universalism; **(.53, .22, -.11)**, *[.45, .20, -.07]*
	Life satisfaction, Emot. stability, Depression-; **(.25, .38, .11)**, *[.48, .55, .40]*
	Wholesome act. int., Intellectual int., Openness; (.31, .21, .0), [.41, .44, .17]∗
	Agreeableness, Empathy, Extraversion; (.27, .32, .0), [.45, .39, .16]∗
	Life satisfaction, Emot. stability, Agreeableness; (.25, .26, .0), [.48, .35, .25]∗

∗Identified also as a 2path pos. motif due to overlap in the motif definition with Imb. triad II T.
(-) after the name of a psychological attribute means that it has been reversed.

Strong PPP triads may indicate the presence of a common cause, for instance, because the three variables measure the same underlying psychological construct, which then acts as a latent variable. Unsurprisingly, the relationships among the three constructs measured by the Body Consciousness questionnaire represent one such case. Another such motif is made of Conscientiousness and two integrity measures, Fair-Mindedness and Self-Disclosure, pointing out that they are likely capturing similar psychological dimension. A second possibility that may underlay PPP triads is a positive feedback between the variables, as found in the mutualism model for intelligence.

All six NNN triads involve Schwartz's values, with Tradition being present in five of them. This configuration cannot emerge from a common cause and may suggest a negative feedback loop between the attributes. Still, such an interpretation is formed on conclusions about intraindividual differences that are based on interindividual data, which may not necessarily hold. A second possible reason for observing NNN triads is that the variables have been conditioned on a common effect to which each of them positively contributes. The logic here is the following. Suppose that three variables A, B, and C increase the probability of common effect D. If we condition on D, we only consider the values of A, B, and C for a given value of D. Suppose we observe that the effect is present (or D has a high value), but A is not present (or has a low value). Then that information makes it more likely that B or C are present (or have a high value). Thus, conditioning on D, we expect A, B, and C to be negatively related so that they form an NNN triangle in the partial correlation network.

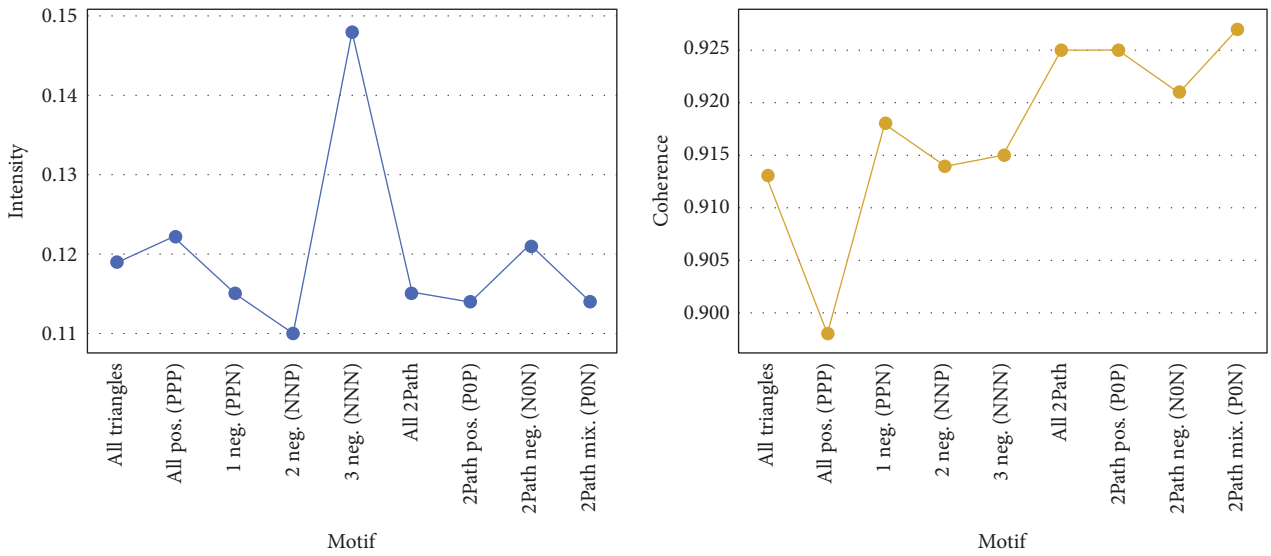


Figure 12: Means of intensity and coherence of all triads and signed motifs in the network.

One NNP triad consists of a negative association between Low Militaristic values and Interests in wholesome activities, while both variables are positively correlated with Universalism. This triad identifies a puzzling relationship that might suggest multidimensionality of the Universalism value. Positive 2paths show that Empathy, Emotional Stability, and Intellectual Interests may play the role of mediators. Negative and mixed 2paths similarly show the variable in central position (position "*J*" in Table 3) as bridging the remaining two attributes in the subgraph. Finally, eight configurations present the strongest imbalanced triplets II T. in the network, which are not possible to describe in the metric space. Three of them also fall under 2paths, due to the overlap in the motif definition. The variable in position "*J*" (see Table 3, first row) in this motif is likely to be a broad concept with multiple meanings.

4.3.2. Analysis of Motif Intensity. In previous research, the Intensity measure has been applied for triadic motifs consisting of positive weights only. Therefore, we modified the approach described by Onnela et al. [53] by calculating I and Q separately for triads with a different configuration of positive and negative ties to allow comparing the Intensities across different motifs. The average Intensity and Coherence for all investigated motifs are shown in Figure 12.

Visual inspection of Figure 12 reveals that the differences in Intensity and Coherence between the motifs are very small (y axes show range of 0.05 for I, and 0.025 for Q). When looking at the structural motifs concerned only about presence and absence of ties, and not their weights, all triads have a higher Intensity than 2paths, but the difference is very small. In psychological networks, it would be expected that triangles have a higher Intensity than 2paths, as triangles represent mutual connections between all three nodes, making it more likely that the nodes will reinforce each other. Because of this reinforcement, it would be expected that the weights are of higher absolute value than in 2paths, where one edge is missing, making such effect less plausible.

The most intensive motif, that is, the motif with the highest average geometric mean of weights, is a triad made of three negative ties NNN, followed by positive triad (PPP) and 2path with two negative ties (N0N). The finding that a NNN motif is the most intensive is somewhat surprising for networks of this kind, but, before attempting interpretation, we will proceed first with analysis of Coherence, followed by significance testing.

Internal Coherence of 2paths (open triads) is somewhat higher than for closed triads (Figure 12, right panel), which is to be expected as 2paths consist of one weight less than triads. PPN seems to have relatively higher, while PPP relatively lower Q.

Having a high (low) average Intensity of a motif does not imply that the motif is highly (lowly) expressed in the network. Therefore, the next step is to check how significant the Intensities are. The same applies to the Q, where a high Q of a motif does not imply it is significantly more coherent. To answer those questions, the Intensities and Coherences of each motif are compared with the mean of I and Q of each motif in an ensemble of 1000 random networks. The results of the analysis are shown in Figure 13.

The only motif whose Intensity (percentile value > 97.5) is significantly high is a triad with three negative ties (NNN), which is in line with the results on the frequency and the descriptive analysis presented in Figure 12. Although the average Intensity is not high in absolute terms (slightly above 0.14), the frequency and Intensity analysis both suggest that the NNN motif is an important characteristic of the network. In Table 3, we saw that all NNNs involve only Schwartz's values. NNN motifs show a tendency to be "nested" around few nodes; only the nodes that represent Schwartz's values are "responsible" for the high frequency (and Intensity) of that

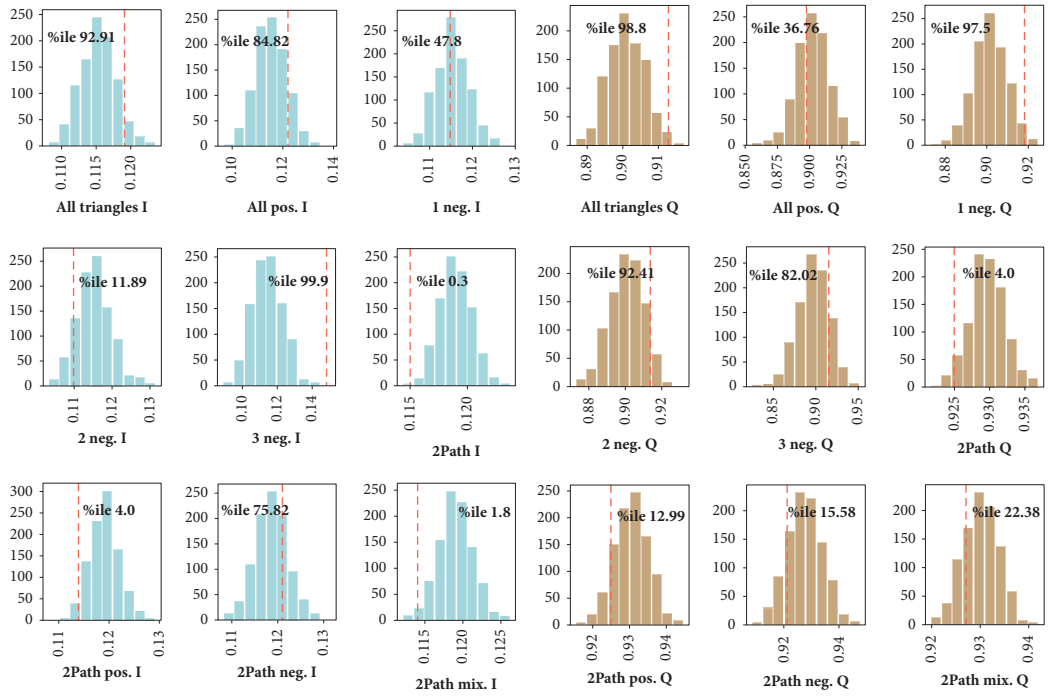


FIGURE 13: Significance of motifs' intensity and coherence: distributions of intensity (first three columns, colored blue) and coherence (last three columns, colored light brown) of all closed and open triads, 2paths, and signed motifs in 1000 random networks with same structure and weight distribution as empirical network, with its percentile values.

motif on a network level. Furthermore, from Figure 1 (and the centrality analyses) we observed that not all Schwartz's values are central. From that we may generate a hypothesis that the most prominent characteristic of the psychological system of 26 attributes is described by a negative feedback between values, although the cluster with such pattern is not central in the system. A second possibility is that some of the values are involved in a common effect with respect to one of them, which might for instance arise when, say, Tradition is caused by all other variables. Due to the conditioning on the common effect, the NNN pattern may arise for the causal variables in the partial correlation network. A final possibility would be that the high occurrence of NNN may be the result of estimating network on a sample which is self-selected (i.e., implicitly conditioned) on a variable that is a common effect of Schwartz's values.

Two motifs with significantly small Intensity (percentile value < 2.5) are all 2paths motif (structural, disregarding the signs of ties), and 2paths with one negative and one positive tie (with mixed ties, P0N). The later finding is an example of the importance of comparison with the reference system. When we analyzed only the average Intensity, we have found that 2paths have a higher Intensity than other motifs. Comparing this to what may be expected given the network structure and weight distribution, we can see that, in fact, the Intensity of 2paths, although somewhat higher in absolute value than Intensity of other motifs, is significantly smaller than it would be expected by the null model. The "intuitive" expectation about smaller Intensity of 2paths due to the lack of third link is supported.

Closed triads (all triangles) display significantly high internal Coherence. From the tie's perspective, this may suggest that weights of similar strengths show the tendency to form triads, or, from a node perspective, that psychological attributes that form a triad tend to be connected with ties of similar strengths (in absolute values). Imbalanced triad (PPN, called "1 neg." in Figure 13) is also significantly more coherent, meaning that the weights within this triangle tend to be equally distributed (they do not show big variations). Interestingly, so-called imbalanced triads in this network consist of "balanced" edge weights. The overall pattern of results show that a significant I does not imply significance in Q, which highlights that they measure two different aspects of this system.

5. General Discussion and Conclusions

This paper has demonstrated how the use of three metrics taken from network science can enrich our understanding about psychological networks. Given the effort invested in estimating the network structure, it is a missed opportunity not to use the information it entails more fully to gain deeper understanding of estimated network. This "omission" may be understood and partly explained by researchers in the field being preoccupied primarily with network estimation methods [11, 47, 62] and replicability issues [49, 63, 64] that arise from the fact that network structures between variables are considerably more difficult to determine, relative to, for example, internet links or electricity nets; after all, conditional association between variables is not observable,

but must be estimated from data. Appropriately dealing with sampling error in estimating network structures, as well as assessing their robustness, has therefore been the priority in psychological network analysis.

The concise overview of the three methods in terms of hypothesis and research questions and procedure is given in Table 4.

We demonstrated on illustrative dataset how each of the methods proposed here adds new information about the network structure. First, the MST can help us in shedding light on the topological arrangement of psychological attributes in the network. Specifically, in the current example, the MST suggests that Empathy is the most similar to all other traits and plays the role of a "network connector;" it is the most central trait when centrality is based on the network filtered down to its most essential ties. In the network which also includes Big Five traits, it was somewhat surprising to see that Empathy has such important standing. This could be due to the questionnaire used for this trait (the Empathy Quotient) which captures affective and cognitive aspects (see Table 1). The authors [39] of the questionnaire state that the cognitive component of Empathy is closely related with an individual's "Theory of Mind," a cognitive process that allows people to understand others and oneself. It might, thus, be plausible that cognitive processes related to Theory of Mind serve as a central hub in the system. In addition, it is tempting to see the analogy and state that the trait which is seen by some to hold society together may also hold this network of different psychological attributes together. This finding is worth of further attention due to an implicit and misguided notion that Big Five traits are the best representative of psychological differences between individuals. If true, in network terms that would imply that they are expected to be in the top five most central nodes, which is the case only for some of them. In fact, Openness is among peripheral nodes. Nonetheless, further theoretical consideration and research is needed. The MST provided an additional insight into possible clusters of attributes and showed that clusters, that is, different branches of the tree, for the most part do not align with different kinds of psychological variables. For example, Big five traits and Schwartz's values are placed on different branches, suggesting that the grouping of variables is based on specific content rather than "nature" of a psychological variable (e.g., whether it is a trait, value, or interest). Furthermore, we used the fact that MST preserves the information of edge signs to employ it for robustness test of network estimation.

Second, by including information about the participation coefficient based on predefined communities, which also included "communities of one," we highlighted the specific role of some nodes based on their equal importance to the structure of different parts of the network. We found that Intelligence, although weakly connected to other traits, and by all centrality measures quite peripheral, does seem to have an interesting property of being relatively equally associated with all different kinds of nodes in the network. Based on this finding we can hypothesize that cognitive ability relates to personality: not in terms of substantial effect sizes but because it relates at a constant strength to most "parts" of psychological system. In other words, the question about relation between cognitive ability and psychological individual differences could be better answered if instead of looking at the "size" of that influence (operationalized with some statistical measure), researchers refocus their attention on the "broadness" of that influence. This agrees with the suggestion of Salovey and Mayer [25] that, instead of looking at pairwise correlations, a more complex analysis that looks at many connections at once should be preferred. Likewise, network ties of Intelligence seem to imply a different relation with Big Five model than reported in the recent review [65]. When 24 other relevant individual differences (26 minus 2 variables whose connection is under consideration) are controlled for, the strongest tie is not with Openness, but with Agreeableness and Extraversion (both negative and around 0.10).

We used PC together with the Participation Ratio to arrive at more sensible centrality measure, which showed that different centrality indices converge to Extraversion, Emotional Stability, and Empathy as the three most central nodes in this network. Centrality of Extraversion and Emotional Stability would be expected since they are one of the traits that have been recognized as important psychological dimensions and systematically studied from early on in psychological science. Empathy taking the "third place" is somewhat surprising, but, as discussed before, could be related with this trait capturing cognitive processes that are essential and fundamental in many social interactions [66].

Finally, we used motif analysis to research possibly interesting three-node configurations and investigate whether this psychological network "behaves" as a social network regarding its balance of negative and positive ties within a triad, and the results showed this is not the case. We learned that some configurations that are challenging to interpret exist in the network at a higher frequency than would be expected in the reference system; most notably, this was the case for NNN triads. Identification of strong motifs revealed that these triads originate mostly from one group of nodes, Schwartz's values, possibly revealing negative feedback or (implicit) conditioning on a common effect of some or all of the variables. NNN triads are also significantly stronger than expected, but otherwise intensity and coherence do not seem to be related with frequency of motifs.

Methodological Considerations Related with the Reverse Coding of Variables. An important issue related with network modeling of relationships between continuous variables which probably did not receive enough attention so far is the effect of reverse coding of variables on the results of network methods (we are grateful to a reviewer for pointing out this issue). It becomes an even more pressing issue when nodes are aggregations of more complex concepts, not easily described as positive or negative (e.g., some values), or when variables present dimensions which are interpretable on both ends (e.g., emotional stability–neuroticism, extraversion–introversion), and often coded arbitrary. This is the case for many continuous variables in psychology, and probably for all variables in our dataset to some extent. For example, Emotional Stability (ES) is often coded negatively as Neuroticism (N), begging the question what would happen

Table 4: An overview of the three methods.

MINIMUM SPANNING TREE	
Recommended for network	Dense networks and/or networks with many small edges
Important procedure steps	(1) Selecting a distance measure: (i) Gower's distance (ii) Distance inversely proportional to the shared variance (2) Centrality analysis (by inspection or/and with standard centrality measures computed)
Output	MST – the filtered network
Methodological considerations	Distance measures (i) and (ii) will produce different MSTs if a network has negative ties*
Other analytical possibilities	Looking at MST branches as communities Using MST to test the robustness of the network estimation of the most essential edges MST can include distance metric as weight for further analysis
Effect of reverse-coding variables	(i) affected* (ii) not affected
Hypothesis/Research question	RQ: Which node is the most central? Which (overlapping) communities exist in the network?
PARTICIPATION COEFFICIENT *as a corrective*	
Recommended for network	(1) Pre-existing differences in the kinds of nodes (2) Networks with communities

Important procedure steps	If (1) is true:
	(a) Defining the node groups
	(b) Calculating PC
	(c) Choosing the centrality measure to be corrected with PC (optional)
	If (2) is true:
	(a) Data-driven detection of communities
	(b) Calculating PC
	(c) Choosing the centrality measure to be corrected with PC
	(d) Comparing the rank order of chosen centrality measure before and after the correction

Output	PC values for each node (in (1)(b) and (2)(b); The corrected centrality measure (for (1)(c) and (2)(c)) Communities should not overlap ((1) and (2))
Methodological considerations	(1) a Group affiliations may be ambiguous (2) a Decision about appropriate community detection algorithm
Other analytical possibilities	PC version that treats positive and negative edges separately
Effect of reverse-coding variables	Not affected if signs are not taken in the account when calculating PC
Hypothesis/Research question	H_0 = A centrality measure is not affected by (pre-existing or data-driven) communities

Table 4: Continued.

MOTIF ANALYSIS	
Recommended for network	(1) Not for networks with small number of nodes and/or very low density (2) Has negative and positive ties (3) If weighted, additional steps in the procedure
Important procedure steps	(a) Defining motifs of interest and the null model (b) Motif identification and frequency (c) Significance testing of motif frequency If (3) is true: (d) Motif intensity (and coherence) (e) Significance testing of motif intensity (and coherence)
Output	Identified motifs; Motif Frequency; P-values for frequencies; Motif intensity (and coherence); P-values for the intensities (and coherence)
Methodological considerations	The definition of the null (reference) model
Other analytical possibilities	Other motif structures, e.g. that include more nodes
Effect of reverse-coding variables	Identified motifs and motif frequency will be different*, but onclusions about significance (of frequency, intensity, and coherence) will tend to converge
Hypothesis/Research question	Many research questions and hypotheses possible. In this study: *Signed edges will tend to cluster in the line with what is observed in social networks and correlational networks (balance theory, forbidden triads, imbalanced triplets):* H_1 = P0P, NNN, and PPN will be less frequent than expected by chance. RQ1 = Do same pattern of results holds true when only relatively stronger motifs are considered? RQ2= Are Intensity and Coherence measure following the same pattern of results as the frequency of motifs?

*This should be the case, but there is a possibility that distances related to negative edges are present in a network in such a way that will not affect the MST construction, e.g. a weak negative tie that exists among two peripheral nodes that have ties to other more central node.

with the results of analyses if we used N instead of ES? To find out we repeated most of the analyses reported in this paper with the network that had N instead of ES, and several other networks with some of the variables recoded. The results are presented in detail in SM (Section 12), while here we will highlight just the most important conclusions. The estimated network will have the same structure and absolute values of weights, but all the edges of reversed node will change their sign. Weight distribution of network is affected too, due to the changes in signs of some of the weights. The most affected are the results of MST, but only if the preferred distance measure is used. Otherwise, with the measure inversely proportional to shared variance, MST results are unaffected. This situation brings up the dilemma of which distance metric to use: the more rigorous one that is affected by variable coding, or the one which leads to a possibly substantial loss of information but is immune to reverse coding? We do not provide an answer, because, as usual, it will depend on the specific network, variables included, and the research question. Nevertheless, researchers need to be aware of this issue. In contrast with MST, PC that takes only absolute value of weights is not affected by reverse coding. Motif analysis will produce different motif frequencies, intensity, and coherence values, but the results of significance testing will not be affected to a greater extent and will tend to converge for the same network with differently coding some of the variables.

A logical conclusion following from previous section is that the three methods discussed in this paper require an effort to be applied to a psychological network, as some additional decisions need to be reached such that they are in accordance with research questions/goals (also explained in Sections 4.1, 4.2, and 4.3). Each decision has its repercussions. In case of MST, one needs to consider the presence of negative ties and what is achieved by deciding to look at two negatively associated nodes as more dissimilar than two nodes that are not connected at all. For PC, the nature of nodes included in the network needs to be carefully looked at, while for motif analysis some notion about which specific configurations may reveal interesting patterns in the network should be formed. The common ground of all three methods is that they look at direct, local ties, but in the contrast to the degree centrality they provide more fine-grained information. This presents a potential for a deeper understanding of any network but is also a very convenient feature for networks that do not have well-defined boundaries. By boundaries, we refer to two issues. The first issue is the possibility that some node(s) which are part of the system are not included in the network analysis. This is an issue for our network where selection of variables was atheoretical, since a "global" theory that describes all psychological attributes does not exist. The selection was further constrained by data availability. For example, we can think of some potentially important attributes that are not in the network, for example, self-efficacy, need for cognition, and narcissism. While acknowledging this, the limitation had its advantage in indirectly preselecting some of the currently most studied/used (and therefore, it could be argued, important) concepts. The second issue is related to the first one and refers to the nature of the investigated network. Some networks are more easily influenced by "externalities;" for example, for a psychological network this may include some important life events that can bring about the change in the network by directly or indirectly influencing one or more nodes. Hence, global properties of such network, and measures relying on all ties in it, may be less useful. The fact that whole system is not represented and that it is an "open" system, as is the case in probably many psychological networks studied so far, was the motivation for introducing these three network methods that rely more heavily on local than global network structure.

To conclude, the added value of more information provided by more complex network tools comes at the price of less straightforward procedures, and making more decisions (hopefully informed by theory and previous research). However, we believe that those elements are just more salient when using these three methods, than when using typical centrality analysis based on different centrality indices, where many assumptions are implicit (e.g., that all nodes are equally likely to be connected to any other node). Therefore, we look at this requirement for higher deliberation as a good practice in general when applying any network analysis to psychometric data, as it challenges researchers to think more about nature of nodes, ties, and smaller network configurations in the network. Nevertheless, that is not an easy task. Understanding these "new" methods may be at first somewhat less straightforward and difficult for researchers not heavily involved in network analysis. This is especially true for motif analysis, which is by far the most complex of the three. Given that network approach is relatively new in psychology, it will take some time for network ideas and methods to "sink in." Unfortunately, it also lacks strong theories. Be that as it may, better understanding of its analytical tools and exploratory (and that sometimes means undertheorized) potential will greatly facilitate the development of such theories. William James's argument that "a degree of vagueness can be beneficial to science when attempting new research directions" [67, p.2] nicely captures the point we are trying to make. This holds true not only for network theories, but also for any kind of theories which aim to integrate many small ("local") theories in psychology.

The methodology presented offers interesting possibilities for applications to other areas. For example, it would be informative to see how equally distributed ties are of depression symptoms among different groups of symptoms (e.g., thoughts, physical symptoms, behaviors, and feelings), and which symptoms are most central when that information is taken into account. We are not suggesting that all methods should be used in every analysis. The most appropriate methods and its specific procedure should be established based on careful consideration of the data at hand, research questions and theory behind it, and knowledge of existing network science tools. Our goal was to expand the latter.

Network approach is often compared to other multivariate methods more commonly used in the field of psychology, for example, structural equation modelling (SEM), confirmatory factor analysis (CFA), mediation analysis (MA), hierarchical clustering (HC), and multidimensional scaling (MDS). Although detailed comparison is out of scope of this

paper, we will proceed with a general overview with highlight on three most notable differences between network approach and most of multivariate methods used in psychology that are more closely related with three specific methods we introduced in this paper. Firstly, the network approach is less directly guided by researcher's assumptions about the connections between variables than most other methods (e.g., CFA), that is, except for the decision about the variables that will be included in the network. In reality, the decision about which variables will be included in the network is constrained with data availability. In this regard, using PC can help in indirectly controlling for some aspects of that constraint, acting as a corrective measure for possible bias in the selection of nodes that have been included in the network.

Secondly, in comparison with SEM, and MA, network analysis usually deals with a greater number of variables at once, implying that SEM and MA may be more appropriate for smaller set of variables, especially if clear theoretical expectations exist about relationships between the constructs.

Finally, other approaches are not trying to look at the set of investigated variables as a system and reveal the properties of that system; they rarely go beyond the microlevel of examining specific connections. In that sense, MST and motif analysis are valuable tools within network approach. MST can be used, among other reasons (mentioned in this paper), to filter the most important connections in the system and to provide answer about the most central variables/nodes on a more general level than specific centrality measures. One part of the output of motif analysis, the identification of motifs, can be viewed as a counterpart to MA (or SEM if configurations tested with motif analysis include more than three variables/nodes) among network methods. However, other outputs of motif analysis, significance of motif frequency, intensity/coherence analysis, and its corresponding significance testing aim at insights that use aggregated information about microlevel to inform about the properties of system as a whole.

In conclusion, at this rather early stage of its application in the field of psychology, network analysis is mostly an exploratory approach, but that is likely to change with the introduction of more sophisticated methods that may provide additional insights. In turn, this will enhance the development of specific network theories that can be explicitly tested, resulting in unique contributions to our knowledge about psychological phenomena.

If we view network approach as a different way of thinking about psychological constructs, then exploring networks more "deeply" may lead us to interesting and important findings that would otherwise be missed. Those findings can lead to new questions, generate new specific hypotheses, and help form truly progressive network theories of psychological phenomena.

Limitations of This Study. Our goal was to demonstrate three methods by applying them to an illustrative dataset. The dataset, however, has some limitations that are important to note. Although we had an atypically large sample (for psychological research), it featured considerable amount of missing data, and how exactly to deal with this problem in network modelling is still an open question [68]. Another open issue in psychological networks is measurement error, which is not accounted for. On an interpretative level, since nodes in network are entities, it is not clear whether their associations can be interpreted as conceptual overlap. To the list of open questions that fall beyond the scope of this paper, we may add the common method variance, which could be responsible for observing some of the edges. However, given that we used partial correlations in the network construction, we believe that most of common method variance (except those unique to a pair of variables) is in that way excluded. Furthermore, one of the sources of common method variance, social desirability, is explicitly included in our network because Self-Disclosure is used as indicator of proclivity to give socially desirable answers (the higher the trait, the smaller the proclivity). Finally, although we had a relatively big sample (pairwise), we do not know how selection bias may influence the results. The trade-offs of "big data" in general is that, on the one hand, it provides more diverse and bigger samples, but, on the other hand, self-selection bias can affect results in many different and unexpected ways. This can play out at multiple levels. For instance, FB users may be unrepresentative regarding some of the traits or due to demographics [69], or FB users who used the *myPersonality* application could be, on average, psychologically different. For example, it could be argued that the sample consists of people who are more interested in psychological aspects of reality and in understanding themselves and others when compared to the general population. In line with this possibility, general self-selection may have influenced our findings about the important role of Empathy in the network. Lastly, individuals chose freely to fulfil certain questionnaire(s). Insomuch as the choice was not random, there is always a possibility that individual psychological attributes influenced that choice (e.g., more depressed individuals could be less likely to fill in an intelligence test).

In the context of those limitations, the findings we arrived at while demonstrating three methods are presented as tentative and their value is in generating new and interesting hypotheses. Furthermore, in our tentative interpretations, due to our network made of well-studied and diverse psychological attributes and due to the scope of this article, we just scratched the surface of many more interesting "small" findings (e.g., each identified triad in Table 3 would be a good starting point for discussion and for generating further hypotheses). That being said, harvesting an already existing dataset, which contains information about many psychological attributes of big number of people, repurposing it to demonstrate "new" methods, and, while doing so, addressing some new and some old questions (network of psychological attributes and cognition-personality relationship) present potentially useful exploratory research.

Future Research. Regarding specific questions related to our dataset, future research would benefit from more theoretically guided inclusion of psychological attributes in the network, including different types of intelligence measures that capture more than g-factor. More objective (behavioral)

measures of attributes would enhance the validity of findings. Longitudinal data (within-subjects networks) and data on specific populations (e.g., regarding mental health, age, gender, and culture) would in addition enable answering questions about network dynamics and network structure. Future research can use simulation studies to determine how exactly each of the methods is affected by differences in network density, size, number of groups, structure, weight distribution, etc. This would be especially interesting for MST, as we explicitly mentioned that it could be used to check the robustness of network estimations. We used PC on what we called "predefined communities," but when there are no differences between nature of psychological attributes PC might be used in typical way as well, which starts from empirically determined communities (such example is given in SM, section 13). Likewise, the PC measure can be extended in such a way that one could calculate it for positive and negative links separately. In the motif analysis, we looked only at triads; future work can include higher-order configurations, motifs that involve more than three nodes (e.g., bow tie).

Finally, we selected three network metrics for this article, but there are other measures and techniques that could be fruitfully used in the analysis of psychological networks (e.g., coefficient of intramodule activity, missing link prediction). The message is that network science methodology develops rapidly, and psychologists using network analysis would do well to embrace the possibilities these methods offer in both, analysis and stating new research questions, hypotheses, and even theories.

Data Availability

The data used to support the findings of this study were supplied by David Stillwell and Michal Kosinski under license and so cannot be made freely available. Requests for access to these data should be made to David Stillwell, contact@mypersonality.org.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Srebrenka Letina conceived the idea for the study, asked for the data access, did the data processing, analyses and visualizations, and wrote the paper. Tessa F. Blanken, Denny Borsboom, and Marie K. Deserno edited the text and its structure and provided feedback on the manuscript.

Acknowledgments

We thank David Stillwell and Michal Kosinski for allowing the access to the myPersonality database (myPersonality.org). The work on this paper was partially sponsored by Central European University Foundation, Budapest (CEUBPF). The theses explained herein are representing the author's own ideas but do not necessarily reflect the opinion of CEUBPF. We acknowledge COSTNET (Cost Action CA15109) in funding the short scientific mission which resulted in this work. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement no. 648693). Denny Borsboom is supported by ERC Consolidator Grant no. 647209. We thank Donald Williams for the help in the estimation of nonregularized partial correlation network and Tamer Khraisha for advice on coding and visualizations.

Supplementary Materials

More details about procedures and results of the analyses are organized in 13 sections of the Supplementary Materials: data processing, sample description, description of missing data, descriptive statistics of 26 psychological attributes, the choice of the estimation method, robustness analyses, network of 26 psychological attributes, analysis of network ties, centrality analysis, correlations between four centrality measures in full network and in MST, the MST with different distance measures, the effect of reverse coding on the analyses, and participation coefficient based on empirical (data-driven) communities. *(Supplementary Materials)*

References

[1] H. L. J. Van Der Maas, C. V. Dolan, R. P. P. P. Grasman, J. M. Wicherts, H. M. Huizenga, and M. E. J. Raijmakers, "A dynamical model of general intelligence: The positive manifold of intelligence by mutualism," *Psychological Review*, vol. 113, no. 4, pp. 842–861, 2006.

[2] A. O. J. Cramer, C. D. Van Borkulo, E. J. Giltay et al., "Major depression as a complex dynamic system," *PLoS ONE*, vol. 11, no. 12, article e0167490, 2016.

[3] G. Costantini, J. Richetin, E. Preti, E. Casini, S. Epskamp, and M. Perugini, "Stability and variability of personality networks. A tutorial on recent developments in network psychometrics," *Personality and Individual Differences*, vol. 136, pp. 68–78, 2017.

[4] J. Dalege, D. Borsboom, F. van Harreveld, and H. L. J. van der Maas, "Network analysis on attitudes: a brief tutorial," *Social Psychological and Personality Science*, vol. 8, no. 5, pp. 528–537, 2017.

[5] V. D. Schmittmann, A. O. J. Cramer, L. J. Waldorp, S. Epskamp, R. A. Kievit, and D. Borsboom, "Deconstructing the construct: A network perspective on psychological phenomena," *New Ideas in Psychology*, vol. 31, no. 1, pp. 43–53, 2013.

[6] A. O. J. Cramer, S. van der Sluis, A. Noordhof et al., "Dimensions of normal personality as networks in search of equilibrium: you can't like parties if you don't like people: dimensions of normal personality as networks," *European Journal of Personality*, vol. 26, no. 4, pp. 414–431, 2012.

[7] D. Borsboom and A. O. J. Cramer, "Network analysis: An integrative approach to the structure of psychopathology," *Annual Review of Clinical Psychology*, vol. 9, pp. 91–121, 2013.

[8] J. Kossakowski and A. O. J. Cramer, "Complex dynamical systems in psychology," in *Network Science in Cognitive Psychology*,

M. Vitevitch, Ed., chapter: Complex Dynamical Systems in Psychology, Rutledge Taylor & Francis Group, 2017.

[9] S. Epskamp, L. J. Waldorp, R. Mõttus, and D. Borsboom, "The gaussian graphical model in cross-sectional and time-series data," *Multivariate Behavioral Research*, vol. 53, no. 4, pp. 453–480, 2018.

[10] C. D. van Borkulo, D. Borsboom, S. Epskamp et al., "A new method for constructing networks from binary data," *Scientific Reports*, vol. 4, no. 5918, 2015.

[11] S. Epskamp and E. I. Fried, "A tutorial on regularized partial correlation networks," *Psychological Methods*, vol. 23, no. 4, pp. 617–634, 2018.

[12] C. D. van Borkulo, L. Boschloo, J. Kossakowski et al., "Comparing network structures on three aspects: A permutation test," In press.

[13] R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B*, vol. 11, no. 1, pp. 193–197, 1999.

[14] T. F. Blanken, M. K. Deserno, J. Dalege et al., "The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks," *Scientific Reports*, vol. 8, no. 1, 2018.

[15] P. J. Jones, A. Heeren, and R. J. McNally, "Commentary: A network theory of mental disorders," *Frontiers in Psychology*, vol. 8, Article ID 1305, 2017.

[16] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, 2005.

[17] P. W. Holland and S. Leinhardt, "Local structure in social networks," *Sociological Methodology*, vol. 7, pp. 1–45, 1976.

[18] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[19] P. L. Ackerman and E. D. Heggestad, "Intelligence, personality, and interests: Evidence for overlapping traits," *Psychological Bulletin*, vol. 121, no. 2, pp. 219–245, 1997.

[20] A. R. Jensen, *The g factor: The Science of Mental Ability*, Praeger Publishers/Greenwood Publishing Group, Westport, CT, USA, 1998.

[21] A. S. Griffin, L. M. Guillette, and S. D. Healy, "Cognition and personality: an analysis of an emerging field," *Trends in Ecology & Evolution*, vol. 30, no. 4, pp. 207–214, 2015.

[22] H. J. Eysenck, "Personality and intelligence: psychometric and experimental approaches," in *Personality and Intelligence*, R. J. Sternberg and P. Ruzgis, Eds., pp. 221–247, Cambridge University Press, Cambridge, Uk, 1994.

[23] A. J. Berg, G. M. Ingersoll, and R. L. Terry, "Canonical analysis of the MMPI and WAIS in a psychiatric sample.," *Psychological Reports*, vol. 56, no. 1, pp. 115–122, 1985.

[24] R. Howard and M. McKillen, "Extraversion and performance in the perceptual maze test," *Personality and Individual Differences*, vol. 11, no. 4, pp. 391–396, 1990.

[25] P. Salovey and J. Mayer, "Some final thoughts about personality and intelligence," in *Personality and Intelligence*, R. J. Sternberg and P. Ruzgis, Eds., pp. 221–247, Cambridge University Press, Cambridge, UK, 1994.

[26] J. F. Kihlstrom and N. Cantor, "Social Intelligence," in *Handbook of intelligence*, R. J. Sterrnberg and S. B. Kaufman, Eds., pp. 564–581, Cambridge University Press, Cambridge, UK, 2011.

[27] D. H. Ford, *Humans as Self-Constructing Living Systems: A Developmental Perspective on Behavior and Personality*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, USA, 1987.

[28] S. D. Smirnov, "Intelligence and personality in the psychological theory of activity," in *Personality and Intelligence*, R. J. Sternberg and P. Ruzgis, Eds., pp. 221–247, Cambridge University Press, Cambridge, UK, 1994.

[29] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *American Psychologist*, vol. 70, no. 6, pp. 543–556, 2015.

[30] D. J. Stillwell and M. Kosinski, "MyPersonality project: example of successful utilization of online social networks for large-scale social research," in *Proceedings of the 1st ACM Workshop on Mobile Systems for Computational Social Science (MobiSys)*, 2012, https://www.gsb.stanford.edu/sites/gsb/files/conf-presentations/stillwell_and_kosinski_2012.pdf.

[31] S. H. Schwartz, "An overview of the schwartz theory of basic values," *Online Readings in Psychology and Culture*, vol. 2, no. 2, 2012.

[32] L. R. Goldberg, J. A. Johnson, H. W. Eber et al., "The international personality item pool and the future of public-domain personality measures," *Journal of Research in Personality*, vol. 40, no. 1, pp. 84–96, 2006.

[33] V. Egan and V. Campbell, "Sensational interests, sustaining fantasies and personality predict physical aggression," *Personality and Individual Differences*, vol. 47, no. 5, pp. 464–469, 2009.

[34] L. C. Miller, R. Murphy, and A. H. Buss, "Consciousness of body: private and public," *Journal of Personality and Social Psychology*, vol. 41, no. 2, pp. 397–406, 1981.

[35] J. Rust, "The validation of the orpheus minor scales in a working population," *Social Behavior and Personality*, vol. 26, no. 4, pp. 399–406, 1998.

[36] D. Watson and J. W. Pennebaker, "Health complaints, stress, and distress: exploring the central role of negative affectivity," *Psychological Review*, vol. 96, no. 2, pp. 234–254, 1989.

[37] M. Snyder, "Self-monitoring of expressive behavior," *Journal of Personality and Social Psychology*, vol. 30, no. 4, pp. 526–537, 1974.

[38] P. M. Lewinsohn, J. R. Seeley, R. E. Roberts, and N. B. Allen, "Center for epidemiologic studies depression scale (CES-D) as a screening instrument for depression among community-residing older adults," *Psychology and Aging*, vol. 12, no. 2, pp. 277–287, 1997.

[39] S. Baron-Cohen and S. Wheelwright, "The empathy quotient: an investigation of adults with asperger syndrome or high functioning autism, and normal sex differences," *Journal of Autism and Developmental Disorders*, vol. 34, no. 2, pp. 163–175, 2004.

[40] W. Pavot and E. Diener, "Review of the satisfaction with life scale," in *Assessing Well-Being*, E. Diener, Ed., vol. 39 of *Social Indicators Research Series*, pp. 101–117, Springer, Dordrecht, Netherlands, 2009.

[41] J. Raven, "The Raven's progressive matrices: change and stability over culture and time," *Cognitive Psychology*, vol. 41, no. 1, pp. 1–48, 2000.

[42] A. Mohammadi and E. C. Wit, "Bayesian Structure Learning in Sparse Gaussian Graphical Models," *Bayesian Analysis*, vol. 10, no. 1, pp. 109–138, 2015.

[43] S. Epskamp, A. O. J. Cramer, L. J. Waldorp, V. D. Schmittmann, and D. Borsboom, "Qgraph: Network visualizations of relationships in psychometric data," *Journal of Statistical Software*, vol. 48, no. 4, pp. 1–18, 2012.

[44] P. J. Jones, "Package "Networktools"," 2018, https://cran.r-project.org/web/packages/networktools/networktools.pdf.

[45] "NetworkX — NetworkX [Internet]," 2018, https://networkx.github.io/.

[46] V. D. Schmittmann, S. Jahfari, D. Borsboom, A. O. Savi, and L. J. Waldorp, "Making large-scale networks from fMRI data," *PLoS ONE*, vol. 10, no. 9, article e0129074, 2015.

[47] D. R. Williams, M. Rhemtulla, A. Wysocki, and P. Rast, "On non-regularized estimation of psychological networks," *PsyArXiv [Internet]*, 2018, https://psyarxiv.com/xr2vf/.

[48] D. R. Williams and P. Rast, "Back to the basics: rethinking partial correlation network methodology," *Open Sci Framew [Internet]*, 2018, https://osf.io/fndru/.

[49] M. K. Forbes, A. G. C. Wright, K. E. Markon, and R. F. Krueger, "Evidence that psychopathology symptom networks have limited replicability," *Journal of Abnormal Psychology*, vol. 126, no. 7, pp. 969–988, 2017.

[50] D. J. Robinaugh, A. J. Millner, and R. J. McNally, "Identifying highly influential nodes in the complicated grief network," *Journal of Abnormal Psychology*, vol. 125, no. 6, pp. 747–757, 2016.

[51] J. C. Gower, "Some distance properties of latent root and vector methods used in multivariate analysis," *Biometrika*, vol. 53, no. 3-4, pp. 325–338, 1966.

[52] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Networks*, vol. 32, no. 3, pp. 245–251, 2010.

[53] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, "Intensity and coherence of motifs in weighted complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 71, no. 6, 2005.

[54] J. Siltaloppi and S. L. Vargo, "Triads: A review and analytical framework," *Marketing Theory*, vol. 17, no. 4, pp. 395–414, 2017.

[55] Y. Kalish and G. Robins, "Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure," *Social Networks*, vol. 28, no. 1, pp. 56–84, 2006.

[56] M. Newman, *Networks: An Introduction*, Oxford University Press, New York, NY, USA, 2010.

[57] M. A. Serrano, M. Boguñá, and R. Pastor-Satorras, "Correlations in weighted networks," *Physical Review E*, vol. 74, no. 5, Article ID 055101, pp. 1–4, 2006.

[58] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory," *Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.

[59] J. A. Davis, "Structural balance, mechanical solidarity, and interpersonal relations," *American Journal of Sociology*, vol. 68, no. 4, pp. 444–462, 1963.

[60] F. Heider, "Attitudes and cognitive organization," *The Journal of Psychology: Interdisciplinary and Applied*, vol. 21, no. 1, pp. 107–112, 1946.

[61] R. Toivonen, M. Kivelä, J. Saramäki, M. Viinikainen, M. Vanhatalo, and M. Sams, "Networks of emotion concepts," *PLoS ONE*, vol. 7, no. 1, Article ID e28883, 2012.

[62] A. P. Christensen, Y. N. Kenett, T. Aste, P. J. Silvia, and T. R. Kwapil, "Network structure of the wisconsin schizotypy scalesshort forms: examining psychometric network filtering approaches," *Behavior Research Methods*, vol. 50, no. 6, pp. 2531–2550, 2018.

[63] S. Epskamp, D. Borsboom, and E. I. Fried, "Estimating psychological networks and their accuracy: A tutorial paper," *Behavior Research Methods*, vol. 50, no. 1, pp. 195–212, 2018.

[64] D. Borsboom, D. J. Robinaugh, M. Rhemtulla, and A. O. J. Cramer, "Robustness and replicability of psychopathology networks," *World Psychiatry*, vol. 17, no. 2, pp. 143-144, 2018.

[65] L. Stankov, "Low correlations between intelligence and big five personality traits: need to broaden the domain of personality," *Journal of Intelligence*, vol. 6, no. 2, p. 26, 2018.

[66] F. S. Ahmed and L. Stephen Miller, "Executive function mechanisms of theory of mind," *Journal of Autism and Developmental Disorders*, vol. 41, no. 5, pp. 667–678, 2011.

[67] M. A. Brandimonte, N. Bruno, and S. Collina, "Cognition," in *Psychological Concepts: An International Historical Perspective*, P. Pawlik and G. d'Ydewalle, Eds., Psychology Press, Howe, UK, 2006.

[68] S. Epskamp, *Network Psychometrics [Ph.D. Dissertation]*, University of Amsterdam, Amsterdam, Netherlands, 2015.

[69] F. T. McAndrew and H. S. Jeong, "Who does what on Facebook? Age, sex, and relationship status as predictors of Facebook use," *Computers in Human Behavior*, vol. 28, no. 6, pp. 2359–2365, 2012.

Hindawi
Complexity
Volume 2019, Article ID 8379321, 8 pages
https://doi.org/10.1155/2019/8379321

Research Article

Human Sensitivity to Community Structure Is Robust to Topological Variation

Elisabeth A. Karuza,[1,2] **Ari E. Kahn**,[3,4] **and Danielle S. Bassett**[4,5,6,7,8]

[1]*Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA*
[2]*Department of Psychology, The Pennsylvania State University, State College, PA 16801, USA*
[3]*Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104, USA*
[4]*Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA*
[5]*Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[6]*Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, USA*
[7]*Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA*
[8]*Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA*

Correspondence should be addressed to Elisabeth A. Karuza; ekaruza@psu.edu

Received 6 July 2018; Revised 10 December 2018; Accepted 29 January 2019; Published 11 February 2019

Guest Editor: Cynthia Siew

Despite mounting evidence that human learners are sensitive to community structure underpinning temporal sequences, this phenomenon has been studied using an extremely narrow set of network ensembles. The extent to which behavioral signatures of learning are robust to changes in community size and number is the focus of the present work. Here we present adult participants with a continuous stream of novel objects generated by a random walk along graphs of 1, 2, 3, 4, or 6 communities comprised of N = 24, 12, 8, 6, and 4 nodes, respectively. Nodes of the graph correspond to a unique object and edges correspond to their immediate succession in the stream. In short, we find that previously observed processing costs associated with community boundaries persist across an array of graph architectures. These results indicate that statistical learning mechanisms can flexibly accommodate variation in community structure during visual event segmentation.

1. Introduction

Segmentation processes, such those involved in extracting words from continuous speech, are the backbone of much of human learning. Tasks essential to the language learner, such as mapping meaning onto sound or combining words into phrases and sentences, first require some understanding of the constituent parts of language. The parsing of sensory input into discrete units is of equal importance in other domains; for example, the perception of event boundaries in visual sequences has been shown to play a key role in active memory [1, 2]. Foundational work by Saffran and colleagues demonstrated that segmentation in the absence of semantic or acoustic cues to word boundaries is driven by the transition probabilities between syllables [3, 4]. More specifically, they found that the successful extraction of structure was due to the relative *difference* in transition probabilities throughout streams of nonsense syllables, characterized by high probabilities within words and low probabilities between words. This simple statistic has since been linked to parsing behavior in both visual and motor learning tasks, suggesting that sensitivity to transition probabilities, or *statistical learning*, extends beyond a single cognitive domain [5–8].

However, while pairwise predictive relationships are clearly a powerful statistic relevant to learning, they represent only one source of statistical information available to learners. As discussed more thoroughly in [9], tasks that demonstrate sensitivity to the central tendency of a distribution, such as in discriminating segments from a phonetic continuum [10], are also considered examples of statistical learning mechanisms at work. An examination of the full scope of statistical information exploited by the learner is a valuable endeavor,

particularly when considering learning effects that are not solely explained by transition probabilities or distributional regularities. It has been demonstrated, for instance, that segmentation processes can be stymied by changes to stimulus structure such as varying the length of units to be parsed from continuous input [11]. In sum, there is insight to be gained from considering whether and when learners tune into multiple sources (or levels) of statistical information.

Though not typically framed in this way, recent developments in the field of network science have effectively extended statistical learning to encompass sensitivity to more complex information such as the global architecture of the environment (for a recent review see [12]). Evidence indicates that when certain broad-scale topological patterns are present, segmentation effects can be elicited even when transition probabilities have been equated between all pairs of elements [13–15]. In the most commonly used experimental design, nodes of the graph represent individual images and edges of the graph represent transitions from one image to another in time. By design, neighbors of each node are richly interconnected with one another, ensuring that the graph displays community structure. However, because degree (number of incident edges) is identical for each node, transition probabilities are stable. Learners exposed to a continuous stream of images generated by a random walk along such a graph *still show* sensitivity to the boundaries between communities. Building on these findings, new evidence suggests that the presence of community structure within temporal sequences might be a particularly privileged type of regularity. For example, increases in processing speed have been observed for motor sequences generated by random walks along modular relative to lattice graphs with an identical number of nodes, edges, and degree distribution [16].

As we develop and test hypotheses about why community structure might be particularly important for learning, it is necessary to clarify the extent to which this specific sensitivity generalizes to variations in graph topology. While the presence of community structure has been repeatedly linked to changes in processing at event boundaries, this phenomenon is nearly always studied using a very narrow ensemble of graphs with nodes of degree $k = 4$ and communities of $N = 5$ nodes and $E = 8$ edges [13–16]. Here, we expand on the limited set of graph architectures previously used by systematically assessing how variations in the number and size of communities impact learning. We note that while we vary community structure, we are careful to hold constant local statistics commonly associated with learning, such as the degree distribution of nodes within a graph (i.e., variations in pairwise transition probabilities). We ask whether previously reported increases in processing time at community boundaries, indicative of learners' expectations that sequences tend to stay within communities, are affected when properties of those communities change.

As a secondary goal, we probe segmentation effects using images of 3-dimensional, clearly manipulable objects as opposed to the more commonly employed fractals or glyphs (but see [17]). Thus, the present series of experiments aims to raise the ecological validity of standard approaches to studying learners' sensitivity to community structure. In doing so, we offer greater insight into how this sensitivity might relate to real-world contexts. For example, a rich research tradition on event segmentation in natural scenes has focused on how the confluence of top-down and bottom-up processes enables perceivers to determine the boundaries of visually-presented activities [18], with a particular focus on how the segmentation processes relate to the encoding of information in memory [19]. Because statistical learning mechanisms are proposed to operate in information-rich contexts, such as natural scenes, it is essential to demonstrate that they can handle complex sensory input [20]. Making use of manipulable, natural-looking objects, the work presented here is a step toward strengthening links to learning outside the laboratory.

2. Materials and Methods

2.1. Participants. Data were collected from 100 unique participants: 20 per each of the 5 experimental conditions in a between-subjects design. We used Amazon Mechanical Turk, an online marketplace in which adult workers complete tasks in exchange for financial compensation. Participants were paid at a rate of $0.10 per minute. To ensure that participants were attending to the task, they also received a completion bonus of $1.00 as well an additional $1.00 bonus if their performance on an orthogonal cover task exceeded 90% accuracy. Methods adhered to the guidelines and regulations of the Institutional Review Board (IRB) of the University of Pennsylvania, which approved all experimental protocols. Participants communicated informed consent prior to completing the experiment.

2.2. Stimuli. Color images of objects used in this experiment were pulled from the 2nd edition of the Novel Object and Unusual Name (NOUN) Database [21]. Novel objects were employed to reduce the possibility that the degree to which an object was recognizable would influence participants' processing times. To narrow down the full set of objects to the subset used here (Figure 1(a)), we selected from the database the 24 most distinct objects (highest mean distance scores based on the Spatial Arrangement Method [22]) that were considered familiar and nameable by 50% or fewer participants. Of the resulting list, we replaced three objects with slightly lower distance scores because their high degree of symmetry meant that participants would be unable to perform a rotation judgment task (see below).

Once the object images were selected, one unique continuous visual stream of 1400 trials was created for each participant. Streams were generated by first assigning an object to a node, and then by randomly walking along the edges comprising one of 6 graph types (Figure 1(b)). Object-to-node correspondence was randomized across subjects. All graphs consisted of an equal number of nodes ($N = 24$), and although the degree of each node differed by graph type (ranging from $k = 23$ in the fully connected graph to $k = 3$ in the graph consisting of 6 communities), their relative distribution was matched. In other words, within a single graph type, the degree was equated for all nodes,

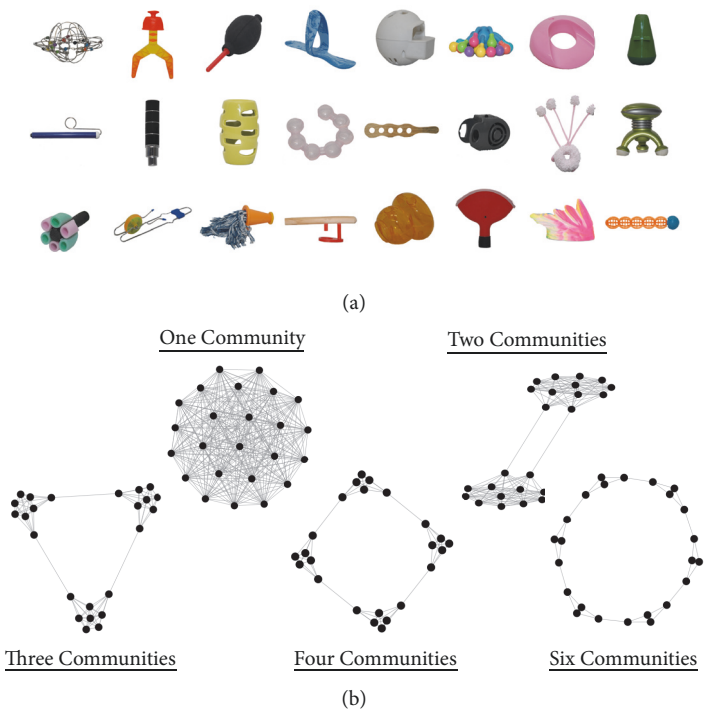


FIGURE 1: Representation of stimulus materials. Panel (a) visualizes the set of NOUN database complex objects assigned to each node in the graphs presented [21]. Panel (b) visualizes the five graph architectures used to generate the continuous object streams. Graph edges correspond to the transition from one object to another in time.

roughly fixing the transition probabilities. Thus, the crucial manipulation was not local variations in pairwise statistics, but rather the number of communities (1, 2, 3, 4, or 6) and the number of nodes within each of those communities (N = 24, 12, 8, 6, and 4, respectively). Because we aimed to maintain dense community structure while also ensuring uniform degree across nodes *within* a graph, the total number of edges differed by graph type, ranging from $E = 276$ in the fully connected graph to $E = 36$ in the graph consisting of 6 communities.

2.3. Procedure. The experimental setup closely mirrored the procedures detailed in [14]; however, for clarity, we summarize our methods here. Participants were instructed to view a continuous stream of objects, and they were informed that over the course of the 35-minute stream, parts of it might become familiar to them. Prior to the initiation of the stream, they were trained to distinguish the canonical orientation of each object from a version that was rotated 90 degrees to the left, and they were tested on their knowledge before moving to the main phase of the experiment. Training trials were repeated until participants achieved an accuracy score of 100% (mean = 83.47 trials, SD = 21.19). The minimum possible number of training trials was 72 (3 trials per object).While viewing the full stream of objects, participants indicated whether each object appeared in its canonical orientation (by pressing 1 on their keyboard) or its rotated version (by pressing 2 on their keyboard). Thus, we were able to collect fine-grained measures of processing time for each object throughout the course of exposure to the stream. From the full set of objects, exactly 15% were rotated. Participants were instructed that they would hear a high-pitched tone if they responded incorrectly during the exposure phase and a low-pitch tone if they responded too slowly. Images of size 300x300 pixels were presented for 1.5 s with no interstimulus interval on a white background.

3. Results

The dependent measure for this experiment was the reaction time (RT) for a canonical, non-rotated image in the stream. Before examining the influence of variation in community structure on this measure, the following steps were taken to clean the data: removal of incorrect or no response trials (7.4% data loss), removal of rotated trials (a further 12% data loss), removal of implausible reaction times (i.e., greater than 1500 ms or less than 100 ms, a further 0.2% data loss), and removal of outlier data points greater than 3 standard deviations from the average RT per subject (a further 1.7% data loss). These preprocessing steps were identical to those used in prior work [14], and we note that the pattern of significant results reported below holds without the removal of implausible and outlier data points. Next, we ran two regression models to answer the following questions: First, do previously reported increases in RTs at community boundaries vary by community size and number (**Model 1**)? Second, are general processing times, separate from the hypothesized cross-community RT increases, influenced by these same

topological variations (**Model 2**)? The linear mixed effects modeling described below was performed with the *lmer()* function (library lme4, v. 1.1–19) in R v. 3.5.1.

3.1. Model 1: Cross-Community Processing Costs. Model 1 was run specifically on data points corresponding to boundary nodes, defined as the nodes directly preceding entry into a new community ("pre-transition nodes") and the nodes representative of that entry ("transition nodes"). Because the fully connected graph contained no boundary nodes, data from this condition were excluded from analysis. We focused specifically on boundary nodes for two reasons: (1) we could not rule out the possibility that learners might show a special sensitivity to boundary nodes regardless of whether they represented entry into a new community; and (2) this approach would ensure a relatively balanced dataset. For instances in which there was a forward and backward traversal of the same cross-community edge (e.g., 24-1-24), we counted only the first pre/transition node pair (24-1). RTs were regressed onto all main effects and interactions of Node Type (pretransition *versus* transition), Community (reverse Helmert coded to test the hypothesis that RTs would increase based on the number of communities) and Trial (continuous from 1–1400, centered to reduce multicollinearity). The model also included the fullest random effects structure that allowed the model to converge: a random intercept for each participant and by-participant random slopes for Trial, Node Type, and their interaction. Results are detailed in Table 1. We observe significant main effects of Node Type (β = 16.79, t = 8.46, and p < 0.001) and Trial (β = -27.35, t = -8.17, and < 0.001). The magnitude of the correlation among fixed effects was less than $r = 0.6$.

To summarize, we find that images associated with transition nodes elicited significantly longer RTs than images occurring directly prior to that transition (Figure 2). As expected, we also find that RTs decreased significantly over time regardless of Node Type; that is, participants overall became faster at making orientation judgments. We observe no main effects of Community, and no interactions with this predictor, suggesting that previously reported cross-community RT increases are robust to fluctuations in community size and number. A subsequent simple effects analysis indicates the effect of Node Type for each level of the Community predictor. Significant effects of Node Type are revealed for the 2-community graph (β = 12.81, t = 2.31, and p = 0.021), the 3-community graph (β = 22.57, t = 5.58, and p < 0.001), the 4-community graph (β = 17.46, t = 5.51, and p < 0.001), and the 6-community graph (β = 14.34, t = 5.88, and p < 0.001). Numerically, the effect of Node Type is weakest for the graph consisting of 2 communities of 12 nodes, but we find no significant difference in cross-community RT increases for this graph relative to the others.

3.1.1. Repetition Priming. Because walks often sampled densely from within a community, there was a higher probability (relative to transition nodes) that a pretransition node would have been viewed in the recent past. To disentangle perceptual priming effects from the top-down expectation that sequences should stay within communities, we followed

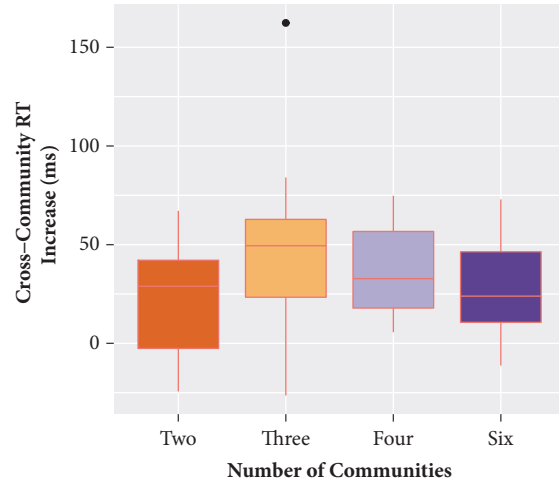


Figure 2: Boxplots (N = 80 participants) illustrating the increase in reaction time across community boundaries for each graph type (2, 3, 4, or 6 communities). Values used to create this plot were calculated by subtracting, for each participant, average RT for nodes *prior* to entry into a new community from average RT for nodes representing entry into a new community. A value of 0 ms would indicate no difference in reaction time at an inter-community boundary.

the approach taken by [14]. Model 1 was rerun with the addition of two confound predictors: Lag10 and Recency. These two predictors indicated the number of times each image was seen in the previous 10 trials and the number of trials elapsed since each image was seen, respectively. Results reveal significant main effects of Lag10 (β = -13.68, t = -8.06, and p < 0.001) and Recency (β = 15.31, t = 9.61, and p < 0.001); however, we maintain our significant main effect of Node Type (β = 5.53, t = 2.66, and p = 0.008). We then subset our data to include only the 30.7% of boundary nodes that had *not* been repeated within the previous 25 trials (any further constraint would have led to an extremely unbalanced dataset). Again, we maintain a significant main effect of Node Type (β = 11.00, t = 2.67, and p = 0.008).

3.2. Model 2: General Processing Times Influenced by Community Size and Number. Procedures for Model 1 were also applied to Model 2. However, as we already confirmed from Model 1 the presence of a processing cost for transition nodes, here in Model 2 we probed RTs for all nodes *except* transition nodes analyzed in Model 1. Because we did not focus exclusively on boundary nodes, we were also able to include data from the fully connected graph. RTs were regressed onto all main effects and interactions of Community and Trial. We again included the fullest random effects structure that allowed the model to converge, which in this case consisted of a random intercept for each participant and by-participant random slopes for Trial. The magnitude of the correlation among fixed effects was less than $r = 0.3$. In addition to the expected main effect of Trial (β = -24.59, t = -11.90, p < 0.001), we also observe a significant main effect of Community for the graph containing 4 communities of $N = 6$ nodes relative to graphs containing 1, 2, and 3 communities (β = -8.83,

Table 1: *Model I* lists the coefficients, t-values, and p-values for each predictor in a model examining the effect of Node Type (pre-transition *versus* transition), Community (reverse-Helmert coded based on graph type), Trial, and their interactions on RTs specific to boundary nodes. *Model II* lists these values for predictors in a model examining the effect of Community, Trial, and the interaction of these predictors for all nontransition RTs. Significant values (corresponding to $p < 0.05$ via the Sattherwaite approximation) are bolded.

Predictor	Coefficient	T-value	P-value
MODEL I			
Community (3 v. 2)	0.76	0.06	0.949
Community (4 v. 3, 2)	-9.31	-1.38	0.172
Community (6 v. 4, 3, 2)	-2.49	-0.53	0.599
Node Type (pre v. transition)	**16.79**	**8.46**	**<0.001**
Trial	**-27.35**	**-8.17**	**<0.001**
Comm (3 v. 2)∗ Node Type	4.88	1.42	0.155
Comm (4 v. 3, 2)∗ Node Type	-0.08	-0.05	0.961
Comm (6 v. 4, 3, 2)∗ Node Type	-0.82	-0.93	0.354
Comm (3 v. 2)∗ Trial	-0.70	-0.14	0.892
Comm (4 v. 3, 2)∗ Trial	-3.02	-1.12	0.267
Comm (6 v. 4, 3, 2)∗ Trial	1.06	0.60	0.553
Node Type ∗ Trial	-0.76	-0.42	0.676
Comm (3 v. 2)∗ Node Type∗Trial	-4.65	-1.45	0.147
Comm (4 v. 3, 2)∗ Node Type∗ Trial	1.23	0.86	0.387
Comm (6 v. 4, 3, 2)∗ Node Type∗ Trial	0.02	0.02	0.984
MODEL II			
Community (2 v. 1)	-9.26	-0.89	0.374
Community (3 v. 2, 1)	-1.63	-0.27	0.787
Community (4 v. 3, 2, 1)	**-8.83**	**-2.09**	**0.040**
Community (6 v. 4, 3, 2, 1)	-1.60	-0.49	0.628
Trial	**-24.59**	**-11.90**	**<0.001**
Community (2 v. 1)∗Trial	1.22	0.37	0.710
Community (3 v. 2, 1)∗ Trial	-1.93	-1.02	0.310
Community (4 v. 3, 2, 1)∗Trial	-1.27	-0.95	0.344
Community (6 v. 4, 3, 2, 1)∗Trial	0.05	0.04	0.965

t = -2.09, p = 0.040; Table 1). Phrased another way, the general processing times after excluding cross-community nodes were most facilitated when learners were presented with sequences generated by a random walk along a graph consisting of 4 communities (Figure 3). Importantly, these effects are observed even when specifically accounting for inter-individual variation in general RTs through the random effects structure of the model. To be clear, when directly comparing general processing times for graphs of 4 communities relative to graphs of 6 communities, we find no significant main effect (β = 9.27, t = 0.84, and p = 0.41). Therefore, it may not be that participants had a particular preference for graphs of 4 communities (of 6 nodes) but that their processing times were generally influenced when information was organized according to many small communities. Finally, to make direct contact with Model 1 analyses, we also reran Model 2 with the inclusion of the Lag10 and Recency predictors described in Section 3.1.1. We note that the main effect of Community (4 v. 3, 2, 1) was marginal (β = -6.98, t = -1.65, p = 0.103). When subsetting to the 18.9% of nodes that had not been repeated within the previous 25 trials, the previously significant main effect of Community (4 v. 3, 2, 1) dropped to β = -4.22, t = -0.97, p = 0.336.

4. Discussion

The current study serves to broaden our understanding of the scope of community-driven learning. We have begun with a replication of prior work demonstrating a processing cost associated with transitioning from one community of objects to another in a continuous sequence [13–16]. In doing so, we have taken the critical step of showing that previously reported effects generalize to novel stimuli that more closely approximate the physical features of manipulable, complex objects found in our real-world environment. The observed increase in reaction time at community boundaries signals that learners are indeed highly sensitive to modular temporal networks; processing costs for transition nodes indicate a violation of the expectation that sequences tend to stay within a given community (for extensive discussion of this point see [14]). The present report also finds no evidence to support the hypothesis that cross-community RT differences are significantly modulated by changes in community size and number. Compellingly, a simple effects analysis pointed to a significant effect of Node Type whether examining 2 communities of N = 12 nodes with degree k = 11 or 6 communities of N = 4 nodes with degree k = 3.

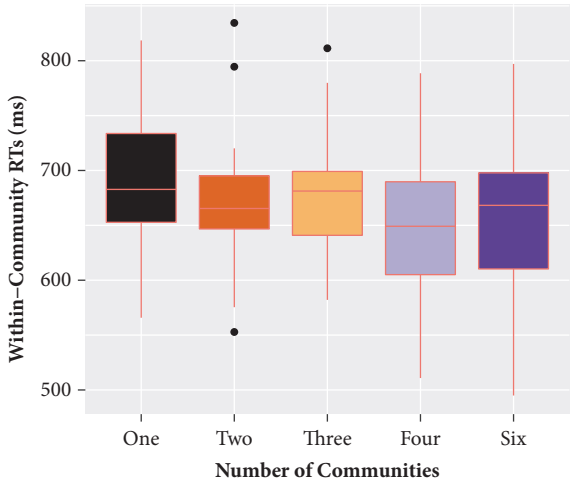


Figure 3: Boxplots (N = 100 participants) illustrating intra-community RTs for each graph type (1, 2, 3, 4, or 6 communities). The greatest general facilitation effects were found for the graph type consisting of 4 communities.

While the present work substantiates the link between community structure and event segmentation (i.e., by focusing on boundary nodes), it is also useful to consider the impact of this property on *general* processing separately from the RT signatures associated with violating learners' expectations that sequences stay within a given community [16]. There is a rich history in cognitive science devoted to uncovering "sweet spots" associated with various cognitive capacities (e.g., [23]). Miller (1956) famously pronounced the limits of verbal working memory as seven plus or minus two [24], and similar constraints are described in tasks of numerical cognition. For example, adults typically have a subitizing range of 5 items beyond which they are unable to automatically determine the number of items in a visual array without counting [25]. While not a constraint *per se*, it could be the case that communities of a certain size or number lead to the most efficient use of processing resources. An analysis of intra-community reaction times (i.e., excluding nodes representative of a transition to a new community) starts to offer an answer to this question. Specifically, results reveal the greatest facilitation of object processing for sequences generated by walks along a graph comprised of 4 communities of N = 6 nodes (compared to graphs with 1, 2, or 3 communities). Given the present design, intended to evaluate segmentation effects while holding constant the total number of nodes and the within-graph degree distribution, it is not possible to disentangle whether the observed facilitation effects are due to the number or size of communities. Nonetheless, this pattern of results has intriguing connections to reports of visual working memory capacity at 4 items [26]. Might it be that community structure is most useful as a cue to underlying structure when the temporal environment is organized into 4 or more groupings? If, however, it is the size of community that affects processing times and not the total number of communities, then we would not find evidence in favor of that hypothesis. Clearly, additional work, using computational as well as behavioral approaches [27], is needed to pinpoint the nature of the relationship between constraints on cognitive capacities and the effects of community structure on sequential object processing.

The sum of these findings confirms that learners are strongly attuned to the presence of community structure and begs further study of the extent to which learning is robust to even more pronounced topological variation. Given that community structure pervades systems as diverse and noisy as linguistic, biological, and social networks [28–30] and that the human brain flexibly accommodates relational information in grid-like maps [17, 31], one would expect learning mechanisms to cope adequately with larger scale and/or sparser instantiations of this property. However, the extent to which learners exploit the full scope of network properties observed in natural systems (e.g., core-periphery structure, scale-free structure, and variations in community-size and density) gives rise to empirical questions to be tested. Specifying the boundary conditions of learning is an especially important area of ongoing and future research. For example, evidence already suggests that local statistics (transition probabilities), when they are sufficiently strong at community boundaries, override typically observed RT increases at event transitions [14]. This tension between local and global regularities, particularly when those regularities are embedded in noisier systems, will be an essential avenue to investigate in greater detail. To follow up on our earlier suggestion that learning in the laboratory should more closely reflect learning outside the laboratory, the focus here on uniform communities and uniform transition probabilities (within-graph) could be considered a limitation given that learners are less likely to encounter such rigidly organized input.

5. Conclusions

We argue here for extending the commonly used definition of statistical learning to encompass learners' sensitivity to the broader topology of their environment. We offer support for this argument by demonstrating that processing times at event boundaries are influenced by temporal community structure across a variety of scales (i.e., the set of graphs tested here), even when object-to-object transition probabilities do not display meaningful variation. Finally, we show that community structure also affects *overall* processing times, with specific facilitative impact observed for sequences comprised of 4 communities of N = 6 nodes relative to fewer communities comprised of a greater number of nodes. The present work marks an important step forward in our understanding of the influence of higher-order architectural properties on learning and processing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors wish to acknowledge Christopher Lynn for helpful comments on this manuscript. This work was supported by the National Science Foundation CAREER award to Danielle S. Bassett (PHY-1554488) and by the Army Research Laboratory through contract number W911NF-10-2-0022. The authors would also like to acknowledge support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Paul G. Allen Foundation, the Army Research Laboratory through contract number W911NF-10-2-0022, the Army Research Office through contract numbers W911NF-14-1-0679 and W911NF-16-1-0474, the National Institute of Health (2-R01-DC-009209-11, 1R01HD086888-01, R01-MH107235, R01-MH107703, R01MH109520, 1R01NS099348, and R21-M MH-106799), the Office of Naval Research, and the National Science Foundation (BCS-1441502, BCS-1631550, and CNS-1626008).

References

[1] K. M. Swallow, J. M. Zacks, and R. A. Abrams, "Event boundaries in perception affect memory encoding and updating," *Journal of Experimental Psychology: General*, vol. 138, no. 2, pp. 236–257, 2009.

[2] C. A. Kurby and J. M. Zacks, "Segmentation in the perception and memory of events," *Trends in Cognitive Sciences*, vol. 12, no. 2, pp. 72–79, 2008.

[3] J. R. Saffran, R. N. Aslin, and E. L. Newport, "Statistical learning by 8-month-old infants," *Science*, vol. 274, no. 5294, pp. 1926–1928, 1996.

[4] J. R. Saffran, E. L. Newport, and R. N. Aslin, "Word segmentation: The role of distributional cues," *Journal of Memory and Language*, vol. 35, no. 4, pp. 606–621, 1996.

[5] J. Fiser and R. N. Aslin, "Encoding multielement scenes: Statistical learning of visual feature hierarchies," *Journal of Experimental Psychology: General*, vol. 134, no. 4, pp. 521–537, 2005.

[6] J. Fiser and R. N. Aslin, "Statistical learning of higher-order temporal structure from visual shape sequences," *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 28, no. 3, pp. 458–467, 2002.

[7] N. Z. Kirkham, J. A. Slemmer, and S. P. Johnson, "Visual statistical learning in infancy: evidence for a domain general learning mechanism," *Cognition*, vol. 83, no. 2, pp. B35–B42, 2002.

[8] R. H. Hunt and R. N. Aslin, "Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners," *Journal of Experimental Psychology: General*, vol. 130, no. 4, pp. 658–680, 2001.

[9] E. D. Thiessen and L. C. Erickson, "Beyond word segmentation," *Current Directions in Psychological Science*, vol. 22, no. 3, pp. 239–243, 2013.

[10] J. Maye, J. F. Werker, and L. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition*, vol. 82, no. 3, pp. B101–B111, 2002.

[11] E. K. Johnson and M. D. Tyler, "Testing the limits of statistical learning for word segmentation," *Developmental Science*, vol. 13, no. 2, pp. 339–345, 2010.

[12] E. A. Karuza, S. L. Thompson-Schill, and D. S. Bassett, "Local patterns to global architectures: influences of network topology on human learning," *Trends in Cognitive Sciences*, vol. 20, no. 8, pp. 629–640, 2016.

[13] S. H. Tompson, A. E. Kahn, E. B. Falk et al., "Individual differences in learning social and non-social network structures," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2017.

[14] E. A. Karuza, A. E. Kahn, S. L. Thompson-Schill, and D. S. Bassett, "Process reveals structure: How a network is traversed mediates expectations about its architecture," *Scientific Reports*, vol. 7, no. 1, 2017.

[15] A. C. Schapiro, T. T. Rogers, N. I. Cordova, N. B. Turk-Browne, and M. M. Botvinick, "Neural representations of events arise from temporal community structure," *Nature Neuroscience*, vol. 16, no. 4, pp. 486–492, 2013.

[16] A. E. Kahn, E. A. Karuza, J. M. Vettel, and D. S. Bassett, "Network constraints on learnability of probabilistic motor sequences," *Nature Human Behaviour*, vol. 2, no. 12, pp. 936–947, 2018.

[17] A. O. Constantinescu, J. X. O'Reilly, and T. E. J. Behrens, "Organizing conceptual knowledge in humans with a gridlike code," *Science*, vol. 352, no. 6292, pp. 1464–1468, 2016.

[18] J. M. Zacks and K. M. Swallow, "Event segmentation," *Current Directions in Psychological Science*, vol. 16, no. 2, pp. 80–84, 2007.

[19] J. M. Zacks, N. K. Speer, J. M. Vettel, and L. L. Jacoby, "Event understanding and memory in healthy aging and dementia of the alzheimer type," *Psychology and Aging*, vol. 21, no. 3, pp. 466–482, 2006.

[20] T. F. Brady and A. Oliva, "Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent: Research article," *Psychological Science*, vol. 19, no. 7, pp. 678–685, 2008.

[21] J. S. Horst and M. C. Hout, "The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research," *Behavior Research Methods*, vol. 48, no. 4, pp. 1393–1409, 2016.

[22] M. C. Hout, S. D. Goldinger, and R. W. Ferguson, "The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling," *Journal of Experimental Psychology: General*, vol. 142, no. 1, pp. 256–281, 2013.

[23] C. Kidd, S. T. Piantadosi, and R. N. Aslin, "The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex," *PLoS ONE*, vol. 7, no. 5, 2012.

[24] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *The Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[25] E. Averbach, "The span of apprehension as a function of exposure duration," *Journal of Verbal Learning and Verbal Behavior*, vol. 2, no. 1, pp. 60–64, 1963.

[26] S. J. Luck and E. K. Vogel, "Visual working memory capacity: From psychophysics and neurobiology to individual differences," *Trends in Cognitive Sciences*, vol. 17, no. 8, pp. 391–400, 2013.

[27] C. Lynn, A. Kahn, and D. Bassett, "Structure from Noise: Mental errors yield abstract representations of events," in *Proceedings of the 2018 Conference on Cognitive Computational Neuroscience*, Philadelphia, Pe, USA, September 2018.

[28] C. S. Q. Siew, "Community structure in the phonological network," *Frontiers in Psychology*, vol. 4, Article ID Article 553, 2013.

[29] M. A. Porter, J.-P. Onnela, and P. J. Mucha, "Communities in networks," *American Mathematical Society*, vol. 56, pp. 0–26, 2009.

[30] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.

[31] M. M. Garvert, R. J. Dolan, and T. E. J. Behrens, "A map of abstract relational knowledge in the human hippocampal–entorhinal cortex," *eLife*, vol. 6, Article ID e17086, 2017.

Hindawi
Complexity
Volume 2019, Article ID 4895891, 39 pages
https://doi.org/10.1155/2019/4895891

Research Article

The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning

R. Harald Baayen,[1] **Yu-Ying Chuang**,[1] **Elnaz Shafaei-Bajestan**,[1] **and James P. Blevins**[2]

[1]*Seminar für Sprachwissenschaft, Eberhard-Karls University of Tübingen, Wilhelmstrasse 19, 72074 Tübingen, Germany*
[2]*Homerton College, University of Cambridge, Hills Road, Cambridge CB2 8PH, UK*

Correspondence should be addressed to R. Harald Baayen; harald.baayen@gmail.com

Received 1 June 2018; Accepted 14 November 2018; Published 1 January 2019

Guest Editor: Dirk Wulff

The discriminative lexicon is introduced as a mathematical and computational model of the mental lexicon. This novel theory is inspired by word and paradigm morphology but operationalizes the concept of proportional analogy using the mathematics of linear algebra. It embraces the discriminative perspective on language, rejecting the idea that words' meanings are compositional in the sense of Frege and Russell and arguing instead that the relation between form and meaning is fundamentally discriminative. The discriminative lexicon also incorporates the insight from machine learning that end-to-end modeling is much more effective than working with a cascade of models targeting individual subtasks. The computational engine at the heart of the discriminative lexicon is linear discriminative learning: simple linear networks are used for mapping form onto meaning and meaning onto form, without requiring the hierarchies of post-Bloomfieldian 'hidden' constructs such as phonemes, morphemes, and stems. We show that this novel model meets the criteria of accuracy (it properly recognizes words and produces words correctly), productivity (the model is remarkably successful in understanding and producing novel complex words), and predictivity (it correctly predicts a wide array of experimental phenomena in lexical processing). The discriminative lexicon does not make use of static representations that are stored in memory and that have to be accessed in comprehension and production. It replaces static representations by states of the cognitive system that arise dynamically as a consequence of external or internal stimuli. The discriminative lexicon brings together visual and auditory comprehension as well as speech production into an integrated dynamic system of coupled linear networks.

1. Introduction

Theories of language and language processing have a long history of taking inspiration from mathematics and computer science. For more than a century, formal logic has been influencing linguistics [1–3], and one of the most widely-known linguistic theories, generative grammar, has strong roots in the mathematics of formal languages [4, 5]. Similarly, Bayesian inference is currently seen as an attractive framework for understanding language processing [6].

However, current advances in machine learning present linguistics with new challenges and opportunities. Methods across machine learning, such as random forests [7, 8] and deep learning [9–11], offer unprecedented prediction accuracy. At the same time, these new approaches confront linguistics with deep fundamental questions, as these models typically are so-called end-to-end models that eschew representations for standard linguistic constructs such as phonemes, morphemes, syllables, and word forms.

There are mainly three possible responses to these new algorithms. A first response is to dismiss machine learning as irrelevant for understanding language and cognition. Given that machine learning algorithms currently outperform algorithms that make use of standard concepts from linguistics, this response is unlikely to be productive in the long run.

A second response is to interpret the units on the hidden layers of deep learning networks as capturing the representations and their hierarchical organization familiar from standard linguistic frameworks. The dangers inherent in this approach are well illustrated by the deep learning model proposed by Hannagan et al. [12] for lexical learning in baboons [13]. The hidden layers of their deep learning network were claimed to correspond to areas in the ventral pathway of the primate brain. However, Scarf et al. [14] reported that pigeons can also learn to discriminate between English words and nonword strings. Given that the avian brain is organized very differently from the primate brain and yet accomplishes the same task, the claim that the layers of Hannagan et al.'s deep learning network correspond to areas in the ventral pathway must be premature. Furthermore, Linke et al. [15] showed that baboon lexical learning can be modeled much more precisely by a two-layer wide learning network. It is also noteworthy that while for vision some form of hierarchical layering of increasingly specific receptive fields is well established [16], it is unclear whether similar neural organization characterizes auditory comprehension and speech production.

A third response is to take the ground-breaking results from machine learning as a reason for rethinking, against the backdrop of linguistic domain knowledge, and at the functional level, the nature of the algorithms that underlie language learning and language processing.

The present study presents the results of an ongoing research program that exemplifies the third kind of response, narrowed down to the lexicon and focusing on those algorithms that play a central role in making possible the basics of visual and auditory comprehension as well as speech production. The model that we propose here brings together several strands of research across theoretical morphology, psychology, and machine learning. Our model can be seen as a computational implementation of paradigmatic analogy in word and paradigm morphology. From psychology, our model inherits the insight that classes and categories are constantly recalibrated as experience unfolds, and that this recalibration can be captured to a considerable extent by very simple principles of error-driven learning. We adopted end-to-end modeling from machine learning, but we kept our networks as simple as possible as the combination of smart features and linear mappings is surprisingly powerful [17, 18]. Anticipating discussion of technical details, we implement linear networks (mathematically, linear mappings) that are based entirely on discrimination as learning mechanism and that work with large numbers of features at much lower levels of representation than in current and classical models.

Section 2 provides the theoretical background for this study and introduces the central ideas of the discriminative lexicon that are implemented in subsequent sections. Section 3 introduces our operationalization of lexical semantics, Section 4 discusses visual and auditory comprehension, and Section 5 presents how we approach the modeling of speech production. This is followed by a brief discussion of how time can be brought into the model (Section 6). In the final section, we discuss the implications of our results.

2. Background

This section lays out some prerequisites that we will build on in the remainder of this study. We first introduce Word and Paradigm morphology, an approach to word structure developed in theoretical morphology within the broader context of linguistics. Word and Paradigm morphology provides the background for our highly critical evaluation of the morpheme as theoretical unit. The many problems associated with morphemes motivated the morpheme-free computational model developed below. Section 2.2 introduces the idea of vector representations for word meanings, as developed within computational linguistics and computer science. Semantic vectors lie at the heart of how we model word meaning. The next Section 2.3 explains how we calculated the semantic vectors that we used in this study. Section 2.4 discusses naive discriminative learning and explains why, when learning the relation between form vectors and semantic vectors, it is advantageous to use linear discriminative learning instead of naive discriminative learning. The last subsection explains how linear discriminative learning provides a mathematical formalization of the notion of proportional analogy that is central to Word and Paradigm morphology.

2.1. Word and Paradigm Morphology. The dominant view of the mental lexicon in psychology and cognitive science is well represented by Zwitserlood [19, p. 583]:

> *Parsing and composition — for which there is ample evidence from many languages — require morphemes to be stored, in addition to information as to how morphemes are combined, or to whole-word representations specifying the combination.*

Words are believed to be built from morphemes, either by rules or by constructional schemata [20], and the meanings of complex words are assumed to be a compositional function of the meanings of their parts.

This perspective on word structure, which has found its way into almost all introductory textbooks on morphology, has its roots in post-Bloomfieldian American structuralism [21]. However, many subsequent studies have found this perspective to be inadequate [22]. Beard [23] pointed out that, in language change, morphological form and morphological meaning follow their own trajectories, and the theoretical construct of the morpheme as a minimal sign combining form and meaning therefore stands in the way of understanding the temporal dynamics of language. Before him, Matthews [24, 25] had pointed out that the inflectional system of Latin is not well served by analyses positing that its fusional system is best analyzed as underlying agglutinative (i.e., as a morphological system in which words consist of sequences of morphemes, as (approximately) in Turkish; see also [26]). Matthews argued that words are the basic units and that proportional analogies between words within paradigms (e.g., *walk*:*walks* = *talk*:*talks*) make the lexicon as a system productive. By positing the word as basic unit, Word and Paradigm morphology avoids a central

problem that confronts morpheme-based theories, namely, that systematicities in form can exist without corresponding systematicities in meaning. Minimal variation in form can serve to distinguish words or to predict patterns of variation elsewhere in a paradigm or class. One striking example is given by the locative cases in Estonian, which express meanings that in English would be realized with locative and directional prepositions. Interestingly, most plural case endings in Estonian are built on the form of the partitive singular (a grammatical case that can mark for instance the direct object). However, the semantics of these plural case forms do not express in any way the semantics of the singular and the partitive. For instance, the form for the partitive singular of the noun for 'leg' is *jalga*. The form for expressing 'on the legs,' *jalgadele*, takes the form of the partitive singular and adds the formatives for plural and the locative case for 'on,' the so-called adessive (see [21, 27], for further details). Matthews [28, p. 92] characterized the adoption by Chomsky and Halle [29] of the morpheme as 'a remarkable tribute to the inertia of ideas,' and the same characterization pertains to the adoption of the morpheme by mainstream psychology and cognitive science. In the present study, we adopt from Word and Paradigm morphology the insight that the word is the basic unit of analysis. Where we take Word and Paradigm morphology a step further is in how we operationalize proportional analogy. As will be laid out in more detail below, we construe analogies as being system-wide rather than constrained to paradigms, and our mathematical formalization better integrates semantic analogies with formal analogies.

2.2. Distributional Semantics. The question that arises at this point is what word meanings are. In this study, we adopt the approach laid out by Landauer and Dumais [30] and approximate word meanings by means of semantic vectors, referred to as embeddings in computational linguistics. Weaver [31] and Firth [32] noted that words with similar distributions tend to have similar meanings. This intuition can be formalized by counting how often words co-occur across documents or within some window of text. In this way, a word's meaning comes to be represented by a vector of reals, and the semantic similarity between two words is evaluated by means of the similarities of their vectors. One such measure is the cosine similarity of the vectors, a related measure is the Pearson correlation of the vectors. Many implementations of the same general idea are available, such as HAL [33], HiDEx [34], and WORD2VEC [35]. Below, we provide further detail on how we estimated semantic vectors.

There are two ways in which semantic vectors can be conceptualized within the context of theories of the mental lexicon. First, semantic vectors could be fixed entities that are stored in memory in a way reminiscent of a standard printed or electronic dictionary, the entries of which consist of a search key (a word's form), and a meaning specification (the information accessed through the search key). This conceptualization is very close to the currently prevalent way of thinking in psycholinguistics, which has adopted a form of naive realism in which word meanings are typically associated with monadic concepts (see, e.g. [36–39]). It is worth noting that when one replaces these monadic concepts by semantic vectors, the general organization of (paper and electronic) dictionaries can still be retained, resulting in research questions addressing the nature of the access keys (morphemes, whole-words, perhaps both), and the process of accessing these keys.

However, there is good reason to believe that word meanings are not fixed, static representations. The literature on categorization indicates that the categories (or classes) that arise as we interact with our environment are constantly recalibrated [40–42]. A particularly eloquent example is given by Marsolek [41]. In a picture naming study, he presented subjects with sequences of two pictures. He asked subjects to say aloud the name of the second picture. The critical manipulation concerned the similarity of the two pictures. When the first picture was very similar to the second (e.g., a grand piano and a table), responses were slower compared to the control condition in which visual similarity was reduced (orange and table). What we see at work here is error-driven learning: when understanding the picture of a grand piano as signifying a grand piano, features such as having a large flat surface are strengthened to the grand piano, and weakened to the table. As a consequence, interpreting the picture of a table has become more difficult, which in turn slows the word naming response.

2.3. Discrimination Learning of Semantic Vectors. What is required, therefore, is a conceptualization of word meanings that allows for the continuous recalibration of these meanings as we interact with the world. To this end, we construct semantic vectors using discrimination learning. We take the sentences from a text corpus and, for each sentence, in the order in which these sentences occur in the corpus, train a linear network to predict the words in that sentence from the words in that sentence. The training of the network is accomplished with a simplified version of the learning rule of Rescorla and Wagner [43], basically the learning rule of Widrow and Hoff [44]. We denote the connection strength from input word i to output word j by w_{ij}. Let δ_i denote an indicator variable that is 1 when input word i is present in sentence t and zero otherwise. Likewise, let δ_j denote whether output word j is present in the sentence. Given a learning rate ρ ($\rho \ll 1$), the change in the connection strength from (input) word i to (output) word j for sentence (corpus time) t, Δ_{ij}, is given by

$$\Delta_{ij} = \delta_i \rho \left(\delta_j - \sum_k \delta_k w_{kj} \right), \tag{1}$$

where δ_i and δ_j vary from sentence to sentence, depending on which cues and outcomes are present in a sentence. Given n distinct words, an $n \times n$ network is updated incrementally in this way. The row vectors of the network's weight matrix $\mathbf{S}$ define words' semantic vectors. Below, we return in more detail to how we derived the semantic vectors used in the present study. Importantly, we now have semantic representations that are not fixed but dynamic: they are constantly recalibrated from sentence to sentence (the property of semantic vectors changing over time as experience unfolds

is not unique to the present approach, but is shared with other algorithms for constructing semantic vectors such as word2vec. However, time-variant semantic vectors are in stark contrast to the monadic units representing meanings in models such as proposed by Baayen et al. [45], Taft [46], Levelt et al. [37]. The distributed representation for words' meanings in the triangle model [47] were derived from WordNet and hence are also time-invariant). Thus, in what follows, a word's meaning is defined as a semantic vector that is a function of time: it is subject to continuous change. By itself, without context, a word's semantic vector at time t reflects its 'entanglement' with other words.

Obviously, it is extremely unlikely that our understanding of the classes and categories that we discriminate between is well approximated just by lexical cooccurrence statistics. However, we will show that text-based semantic vectors are good enough as a first approximation for implementing a computationally tractable discriminative lexicon.

Returning to the learning rule (1), we note that it is well-known to capture important aspects of the dynamics of discrimination learning [48, 49]. For instance, it captures the blocking effect [43, 50, 51] as well as order effects in learning [52]. The blocking effect is the finding that when one feature has been learned to perfection, adding a novel feature does not improve learning. This follows from (1) which is that as w_{ij} tends towards 1, Δ_{ij} tends towards 0. A novel feature, even though by itself it is perfectly predictive, is blocked from developing a strong connection weight of its own. The order effect has been observed for, e.g., the learning of words for colors. Given objects of fixed shape and size but varying color and the corresponding color words, it is essential that the objects are presented to the learner before the color word. This ensures that the object's properties become the features for discriminating between the color words. In this case, application of the learning rule (1) will result in a strong connection between the color features of the object to the appropriate color word. If the order is reversed, the color words become features predicting object properties. In this case, application of (1) will result in weights from a color word to object features that are proportional to the probabilities of the object's features in the training data.

Given the dynamics of discriminative learning, static lexical entries are not useful. Anticipating more detailed discussion below, we argue that actually the whole dictionary metaphor of lexical access is misplaced and that in comprehension meanings are dynamically created from form and that in production forms are dynamically created from meanings. Importantly, what forms and meanings are created will vary from case to case, as all learning is fundamentally incremental and subject to continuous recalibration. In order to make the case that this is actually possible and computationally feasible, we first introduce the framework of naive discriminative learning, discuss its limitations, and then lay out how this approach can be enhanced to meet the goals of this study.

2.4. Naive Discriminative Learning. Naive discriminative learning (NDL) is a computational framework, grounded in learning rule (1), that was inspired by prior work on discrimination learning (e.g., [52, 53]). A model for reading complex words was introduced by Baayen et al. [54]. This study implemented a two-layer linear network with letter pairs as input features (henceforth cues) and units representing meanings as output classes (henceforth outcomes). Following Word and Paradigm morphology, this model does not include form representations for morphemes or word forms. It is an end-to-end model for the relation between form and meaning that sets itself the task to predict word meanings from sublexical orthographic features. Baayen et al. [54] were able to show that the extent to which cues in the visual input support word meanings, as gauged by the activation of the corresponding outcomes in the network, reflects a wide range of phenomena reported in the lexical processing literature, including the effects of frequency of occurrence, family size [55], relative entropy [56], phonaesthemes (nonmorphemic sounds with a consistent semantic contribution such as *gl* in *glimmer*, *glow*, and *glisten*) [57], and morpho-orthographic segmentation [58].

However, the model of Baayen et al. [54] adopted a stark form of naive realism, albeit just for reasons of computational convenience: A word's meaning was defined in terms of the presence or absence of an outcome (δ_j in (1)). Subsequent work sought to address this shortcoming by developing semantic vectors, using learning rule (1) to predict words from words, as explained above [58, 59]. The term "lexome" was introduced as a technical term for the outcomes in a form-to-meaning network, conceptualized as pointers to semantic vectors.

In this approach, lexomes do double duty. In a first network, lexomes are the outcomes that the network is trained to discriminate between given visual input. In a second network, the lexomes are the 'atomic' units in a corpus that serve as the input for building a distributional model with semantic vectors. Within the theory unfolded in this study, the term lexome refers only to the elements from which a semantic vector space is constructed. The dimension of this vector space is equal to the number of lexomes, and each lexome is associated with its own semantic vector.

It turns out, however, that mathematically a network discriminating between lexomes given orthographic cues, as in naive discriminative learning models, is suboptimal. In order to explain why the set-up of NDL is suboptimal, we need some basic concepts and notation from linear algebra, such as matrix multiplication and matrix inversion. Readers unfamiliar with these concepts are referred to Appendix A, which provides an informal introduction.

2.4.1. Limitations of Naive Discriminative Learning. Mathematically, naive discrimination learning works with two matrices, a cue matrix $\mathbf{C}$ that specifies words' form features, and a matrix $\mathbf{S}$ that specifies the targeted lexomes [61]. We illustrate the cue matrix for four words, *aaa, aab, abb*, and *abb* and their letter bigrams *aa, ab,* and *bb*. A cell c_{ij} is 1 if word i has bigram j and zero otherwise.

$$\mathbf{C} = \begin{array}{c} \\ \text{aaa} \\ \text{aab} \\ \text{abb} \\ \text{abb} \end{array} \begin{array}{c} \begin{array}{ccc} \text{aa} & \text{ab} & \text{bb} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{array}. \tag{2}$$

The third and fourth words are homographs: they share exactly the same bigrams. We next define a target matrix **S** that defines for each of the four words the corresponding lexome λ. This is done by setting one bit to 1 and all other bits to zero in the row vectors of this matrix:

$$\mathbf{S} = \begin{array}{c} \\ \text{aaa} \\ \text{aab} \\ \text{abb} \\ \text{abb} \end{array} \begin{array}{c} \begin{array}{cccc} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}. \tag{3}$$

The problem that arises at this point is that the word forms jointly set up a space with three dimensions, whereas the targeted lexomes set up a four-dimensional space. This is problematic for the following reason. A linear mapping of a lower-dimensional space $\mathcal{A}$ onto a higher-dimensional space $\mathcal{B}$ will result in a subspace of $\mathcal{B}$ with a dimensionality that cannot be greater than that of $\mathcal{A}$ [62]. If $\mathcal{A}$ is a space in $\mathbb{R}^2$ and $\mathcal{B}$ is a space in $\mathbb{R}^3$, a linear mapping of $\mathcal{A}$ onto $\mathcal{B}$ will result in a plane in $\mathcal{B}$. All the points in $\mathcal{B}$ that are not on this plane cannot be reached from $\mathcal{A}$ with the linear mapping.

As a consequence, it is impossible for NDL to perfectly discriminate between the four lexomes (which set up a four-dimensional space) given their bigram cues (which jointly define a three-dimensional space). For the input bb, the network therefore splits its support equally over the lexomes λ_3 and λ_4. The transformation matrix **F** is

$$\mathbf{F} = \mathbf{C}'\mathbf{S} = \begin{array}{c} \\ \text{aa} \\ \text{ab} \\ \text{bb} \end{array} \begin{array}{c} \begin{array}{cccc} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0.5 & 0.5 \end{pmatrix} \end{array}, \tag{4}$$

and the matrix with predicted vectors $\widehat{\mathbf{S}}$ is

$$\widehat{\mathbf{S}} = \mathbf{C}\mathbf{F} = \begin{array}{c} \\ \text{aaa} \\ \text{aab} \\ \text{abb} \\ \text{abb} \end{array} \begin{array}{c} \begin{array}{cccc} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix} \end{array}. \tag{5}$$

When a cue matrix **C** and a target matrix **S** are set up for large numbers of words, the dimension of the NDL cue space will be substantially smaller than the space of **S**, the dimension of which is equal to the number of lexomes. Thus, in the set-up of naive discriminative learning, we do take into account that words tend to have similar forms, but we do *not* take into account that words are also similar in meaning. By setting up **S** as completely orthogonal, we are therefore making it unnecessarily hard to relate form to meaning.

This study therefore implements discrimination learning with semantic vectors of reals replacing the one-bit-on row vectors of the target matrix **S**. By doing so, we properly reduce the dimensionality of the target space, which in turn makes discriminative learning more accurate. We will refer to this new implementation of discrimination learning as *linear* discriminative learning (LDL) instead of *naive* discriminative learning, as the outcomes are no longer assumed to be independent and the networks are mathematically equivalent to linear mappings onto continuous target matrices.

2.5. Linear Transformations and Proportional Analogy. A central concept in word and paradigm morphology is that the form of a regular inflected form stands in a relation of proportional analogy to other words in the inflectional system. As explained by Matthews ([25], p. 192f),

> *In effect, we are predicting the inflections of servus by analogy with those of dominus. As Genitive Singular domini is to Nominative Singular dominus, so x (unknown) must be to Nominative Singular servus. What then is x? Answer: it must be servi. In notation, dominus domini = servus servi. ([25], p. 192f)*

Here, form variation is associated with 'morphosyntactic features.' Such features are often naively assumed to function as proxies for a notion of 'grammatical meaning.' However, in word and paradigm morphology, these features actually represent something more like distribution classes. For example, the accusative singular forms of nouns belonging to different declensions will typically differ in form and be identified as the 'same' case in different declensions by virtue of distributional parallels. The similarity in meaning then follows from the distributional hypothesis, which proposes that linguistic items with similar distributions have similar meanings [31, 32]. Thus, the analogy of forms

$$\text{dominus} : \text{domini} = \text{servus} : \text{servi} \tag{6}$$

is paralleled by an analogy of distributions d:

$$d\,(\text{dominus}) : d\,(\text{domini}) = d\,(\text{servus}) : d\,(\text{servi})\,. \tag{7}$$

Borrowing notational conventions of matrix algebra, we can write

$$\begin{pmatrix} \text{dominus} \\ \text{domini} \\ \text{servus} \\ \text{servi} \end{pmatrix} \sim \begin{pmatrix} d\,(\text{dominus}) \\ d\,(\text{domini}) \\ d\,(\text{servus}) \\ d\,(\text{servi}) \end{pmatrix}. \tag{8}$$

In this study, we operationalize the proportional analogy of word and paradigm morphology by replacing the word forms in (8) by vectors of features that are present or absent in a word and by replacing words' distributions by semantic vectors. Linear mappings between form and meaning formalize

two distinct proportional analogies, one analogy for going from form to meaning and a second analogy for going from meaning to form. Consider, for instance, a semantic matrix **S** and a form matrix **C**, with rows representing words:

$$\mathbf{S} = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \qquad \mathbf{C} = \begin{pmatrix} p_1 & p_2 \\ q_1 & q_2 \end{pmatrix} \tag{9}$$

The transformation matrix **G** that maps **S** onto **C**,

$$\underbrace{\frac{1}{a_1 b_2 - a_2 b_1}\begin{pmatrix} b_2 & -a_2 \\ -b_1 & a_1 \end{pmatrix}}_{\mathbf{S}^{-1}} \underbrace{\begin{pmatrix} p_1 & p_2 \\ q_2 & q_2 \end{pmatrix}}_{\mathbf{C}} = \underbrace{\frac{1}{a_1 b_2 - a_2 b_1}\left[\begin{pmatrix} b_2 p_1 & b_2 p_2 \\ -b_1 p_1 & -b_1 p_2 \end{pmatrix} + \begin{pmatrix} -a_2 q_1 & -a_2 q_2 \\ a_1 q_1 & a_1 q_2 \end{pmatrix}\right]}_{\mathbf{G}}, \tag{10}$$

can be written as the sum of two matrices (both of which are scaled by $a_1 b_2 - a_2 b_1$). The first matrix takes the first form vector of **C** and weights it by the elements of the second semantic vector. The second matrix takes the second form vector of **C** and weights this by the elements of the first semantic vector of **S**. Thus, with this linear mapping, a predicted form vector is a semantically weighted mixture of the form vectors of **C**.

The remainder of this paper is structured as follows. We first introduce the semantic vector space **S** that we will be using, which we derived from the TASA corpus [63, 64]. Next, we show that we can model word comprehension by means of a matrix (or network) **F** that transforms the cue row vectors of **C** into the semantic row vectors of **S**, i.e., **CF** = **S**. We then show that we can model the production of word forms given their semantics by a transformation matrix **G**, i.e., **SG** = **C**, where **C** is a matrix specifying for each word which triphones it contains. As the network is informative only about which triphones are at issue for a word but remains silent about their order, an algorithm building on graph theory is used to properly order the triphones. All models are evaluated on their accuracy and are also tested against experimental data.

In the general discussion, we will reflect on the consequences for linguistic theory of our finding that it is indeed possible to model the lexicon and lexical processing using systemic discriminative learning, without requiring 'hidden units' such as morphemes and stems.

3. A Semantic Vector Space Derived from the TASA Corpus

The first part of this section describes how we constructed semantic vectors. Specific to our approach is that we constructed semantic vectors not only for content words, but also for inflectional and derivational functions. The semantic vectors for these functions are of especial importance for the speech production component of the model. The second part of this section addresses the validation of the semantic vectors, for which we used paired associate learning scores, semantic relatedness ratings, and semantic plausibility and semantic transparency ratings for derived words.

3.1. Constructing the Semantic Vector Space. The semantic vector space that we use in this study is derived from the TASA corpus [63, 64]. We worked with 752,130 sentences from this corpus, to a total of 10,719,386 word tokens. We used the `treetagger` software [65] to obtain for each word token its stem and a part of speech tag. Compounds written with a space, such as *apple pie* and *jigsaw puzzle* were, when listed in the CELEX lexical database [66], joined into one onomasiological unit. Information about words' morphological structure was retrieved from CELEX.

In computational morphology, several options have been explored for representing the meaning of affixes. Mitchell and Lapata [67] proposed to model the meaning of an affix as a vector that when added to the vector **b** of a base word gives the vector **d** of the derived word. They estimated this vector by calculating $\mathbf{d}_i - \mathbf{b}_i$ for all available pairs i of base and derived words and taking the average of the resulting vectors. Lazaridou et al. [68] and Marelli and Baroni [60] modeled the semantics of affixes by means of matrix operations that take **b** as input and produce **d** as output, i.e.,

$$\mathbf{d} = \mathbf{M}\mathbf{b}, \tag{11}$$

whereas Cotterell et al. [69] constructed semantic vectors as latent variables in a Gaussian graphical model. What these approaches have in common is that they derive or impute semantic vectors given the semantic vectors of words produced by algorithms such as `word2vec`, algorithms which work with (stemmed) words as input units.

From the perspective of discriminative linguistics, however, it does not make sense to derive one meaning from another. Furthermore, rather than letting writing conventions dictate what units are accepted as input to one's favorite tool for constructing semantic vectors, including *not* and *again* but excluding the prefixes *un-*, *in-*, and *re-*, we included not only lexomes for content words but also lexomes for prefixes and suffixes as units for constructing a semantic vector space. As a result, semantic vectors for prefixes and suffixes are obtained straightforwardly together with semantic vectors for content words.

To illustrate this procedure, consider the sentence *the boys' happiness was great to observe*. Standard methods will apply stemming and remove stop words, resulting in the lexomes BOY, HAPPINESS, GREAT, and OBSERVE being the input to the algorithm constructing semantic vectors from lexical co-occurrences. In our approach, by contrast, the lexomes considered are THE, BOY, HAPPINESS, BE, GREAT, TO, and OBSERVE and in addition PLURAL, NESS, PAST. Stop words are retained (although including function words may seem counterproductive from the perspective of semantic vectors in natural language processing applications, they are retained in the present approach for two reasons. First, since in our model, semantic vectors drive speech production, in order to model the production of function words, semantic

vectors for function words are required. Second, although the semantic vectors of function words typically are less informative (they typically have small association strengths to very large numbers of words); they still show structure, as illustrated in Appendix C for pronouns and prepositions.), inflectional endings are replaced by the inflectional functions they subserve, and for derived words, the semantic function of the derivational suffix is identified as well. In what follows, we outline in more detail what considerations motivate this way of constructing the input for our algorithm constructing semantic vectors. We note here that our method for handling morphologically complex words can be combined with any of the currently available algorithms for building semantic vectors. We also note here that we are not concerned with the question of how lexomes for plurality or tense might be induced from word forms, from world knowledge, or any combination of the two. All we assume is, first, that anyone understanding the sentence *the boys' happiness was great to observe* knows that more than one boy is involved and that the narrator situates the event in the past. Second, we take this understanding to drive learning.

We distinguished the following seven inflectional functions: COMPARATIVE and SUPERLATIVE for adjectives, SINGULAR and PLURAL for nouns, and PAST, PERFECTIVE, CONTINUOUS, PERSISTENCE (denotes the persistent relevance of the predicate in sentences such as *London is a big city*.), and PERSON3 (third person singular) for verbs.

The semantics of derived words tend to be characterized by idiosyncracies [70] that can be accompanied by formal idiosyncracies (e.g., *business* and *resound*) but often are formally unremarkable (e.g., *heady*, with meanings as diverse as intoxicating, exhilerating, impetuous, and prudent). We therefore paired derived words with their own content lexomes, as well as with a lexome for the derivational function expressed in the derived word. We implemented the following derivational lexomes: ORDINAL for ordinal numbers, NOT for negative *un* and *in*, UNDO for reversative *un*, OTHER for nonnegative *in* and its allomorphs, ION for *ation, ution, ition,* and *ion*, EE for *ee*, AGENT for agents with *er*, INSTRUMENT for instruments with *er*, IMPAGENT for impersonal agents with *er*, CAUSER for words with *er* expressing causers (the differentiation of different semantic functions for *er* was based on manual annotation of the list of forms with *er*; the assignment of a given form to a category was not informed by sentence context and hence can be incorrect), AGAIN for *re*, and NESS, ITY, ISM, IST, IC, ABLE, IVE, OUS, IZE, ENCE, FUL, ISH, UNDER, SUB, SELF, OVER, OUT, MIS, and DIS for the corresponding prefixes and suffixes. This set-up of lexomes is informed by the literature on the semantics of English word formation [71–73] and targets affixal semantics irrespective of affixal forms (following [74]).

This way of setting up the lexomes for the semantic vector space model illustrates an important aspect of the present approach, namely, that lexomes target what is understood, and not particular linguistic forms. Lexomes are not units of form. For example, in the study of Geeraert et al. [75], the forms *die, pass away*, and *kick the bucket* are all linked to the same lexome DIE. In the present study, we have not attempted to identify idioms and, likewise, no attempt was made to disambiguate homographs. These are targets for further research.

In the present study, we did implement the classical distinction between inflection and word formation by representing inflected words with a content lexome for their stem but derived words with a content lexome for the derived word itself. In this way, the sentence *scientists believe that exposure to shortwave ultraviolet rays can cause blindness* is associated with the following lexomes: SCIENTIST, PL, BELIEVE, PERSISTENCE, THAT, EXPOSURE, SG, TO, SHORTWAVE, ULTRAVIOLET, RAY, CAN, PRESENT, CAUSE, BLINDNESS, and NESS. In our model, sentences constitute the learning events; i.e., for each sentence, we train the model to predict the lexomes in that sentence from the very same lexomes in that sentence. Sentences, or for spoken corpora, utterances, are more ecologically valid than windows of say two words preceeding and two words following a target word — the interpretation of a word may depend on the company that it keeps at long distances in a sentence. We also did not remove any function words, contrary to standard practice in distributional semantics (see Appendix C for a heatmap illustrating the kind of clustering observable for pronouns and prepositions). The inclusion of semantic vectors for both affixes and function words is necessitated by the general framework of our model, which addresses not only comprehension but also production and which in the case of comprehension needs to move beyond lexical identification, as inflectional functions play an important role during the integration of words in sentence and discourse.

Only words with a frequency exceeding 8 occurrences in the TASA corpus were assigned lexomes. This threshold was imposed in order to avoid excessive data sparseness when constructing the distributional vector space. The number of different lexomes that met the criterion for inclusion was 23,562.

We constructed a semantic vector space by training an NDL network on the TASA corpus, using the **ndl2** package for R [76]. Weights on lexome-to-lexome connections were recalibrated sentence by sentence, in the order in which they appear in the TASA corpus, using the learning rule of NDL (i.e., a simplified version of the learning rule of Rescorla and Wagner [43] that has only two free parameters, the maximum amount of learning λ (set to 1) and a learning rate ρ (set to 0.001). This resulted in a 23,562 × 23,562 matrix. Sentence-based training keeps the carbon footprint of the model down, as the number of learning events is restricted to the number of utterances (approximately 750,000) rather than the number of word tokens (approximately 10,000,000). The row vectors of the resulting matrix, henceforth **S**, are the semantic vectors that we will use in the remainder of this study (work is in progress to further enhance the **S** matrix by including word sense disambiguation and named entity recognition when setting up lexomes. The resulting lexomic version of the TASA corpus, and scripts (in python, as well as in R) for deriving the **S** matrix will be made available at `http://www.sfs.uni-tuebingen.de/~hbaayen/`.).

The weights on the main diagonal of the matrix tend to be high, unsurprisingly, as in each sentence on which the model is trained, each of the words in that sentence is an excellent

predictor of that same word occurring in that sentence. When the focus is on semantic similarity, it is useful to set the main diagonal of this matrix to zero. However, for more general association strengths between words, the diagonal elements are informative and should be retained.

Unlike standard models of distributional semantics, we did not carry out any dimension reduction, using, for instance, singular value decomposition. Because we do not work with latent variables, each dimension of our semantic space is given by a column vector of **S**, and hence each dimension is linked to a specific lexome. Thus, the row vectors of **S** specify, for each of the lexomes, how well this lexome discriminates between all the lexomes known to the model.

Although semantic vectors of length 23,562 can provide good results, vectors of this length can be challenging for statistical evaluation. Fortunately, it turns out that many of the column vectors of **S** are characterized by extremely small variance. Such columns can be removed from the matrix without loss of accuracy. In practice, we have found it suffices to work with approximately 4 to 5 thousand columns, selected calculating column variances and using only those columns with a variance exceeding a threshold value.

3.2. Validation of the Semantic Vector Space. As shown by Baayen et al. [59] and Milin et al. [58, 77], measures based on matrices such as **S** are predictive for behavioral measures such as reaction times in the visual lexical decision task, as well as for self-paced reading latencies. In what follows, we first validate the semantic vectors of **S** on two data sets, one data set with accuracies in paired associate learning, and one dataset with semantic similarity ratings. We then considered specifically the validity of semantic vectors for inflectional and derivational functions, by focusing first on the correlational structure of the pertinent semantic vectors, followed by an examination of the predictivity of the semantic vectors for semantic plausibility and semantic transparency ratings for derived words.

3.2.1. Paired Associate Learning. The paired associate learning (PAL) task is a widely used test in psychology for evaluating learning and memory. Participants are given a list of word pairs to memorize. Subsequently, at testing, they are given the first word and have to recall the second word. The proportion of correctly recalled words is the accuracy measure on the PAL task. Accuracy on the PAL test decreases with age, which has been attributed to cognitive decline over the lifetime. However, Ramscar et al. [78] and Ramscar et al. [79] provide evidence that the test actually measures the accumulation of lexical knowledge. In what follows, we use the data on PAL performance reported by desRosiers and Ivison [80]. We fitted a linear mixed model to accuracy in the PAL task as a function of the Pearson correlation r of paired words' semantic vectors in **S** (but with weights on the main diagonal included and using the 4275 columns with highest variance), with random intercepts for word pairs, sex and age as control variables and crucially, an interaction of r by age. Given the findings of Ramscar and colleagues, we expect to find that the slope of r (which is always negative) as a predictor of PAL accuracy increases with age (indicating decreasing accuracy). Table 1 shows that this prediction is born out. For age group 20–29 (the reference level of age), the slope for r is estimated at 0.31. For the next age level, this slope is adjusted upward by 0.12, and as age increases these upward adjustments likewise increase to 0.21, 0.24, and 0.32 for age groups 40–49, 50-59, and 60-69, respectively. Older participants know their language better and hence are more sensitive to the semantic similarity (or lack thereof) of the words that make up PAL test pairs. For the purposes of the present study, the solid support for r as a predictor for PAL accuracy contributes to validating the row vectors of **S** as semantic vectors.

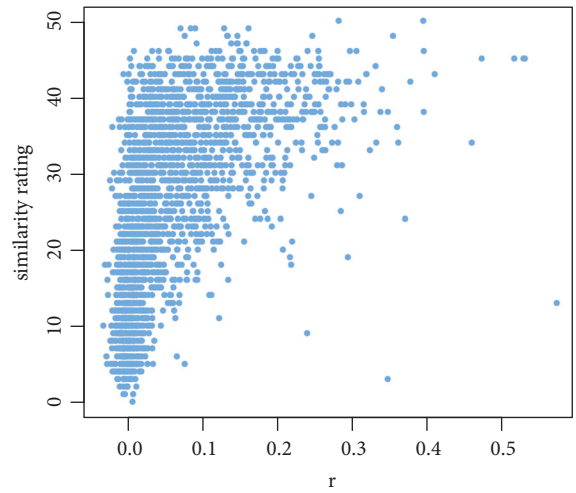


Figure 1: Similarity rating in the MEN dataset as a function of the correlation r between the row vectors in **S** of the words in the test pairs.

3.2.2. Semantic Relatedness Ratings. We also examined performance of **S** on the MEN test collection [81] that provides for 3000 word pairs crowd sourced ratings of semantic similarity. For 2267 word pairs, semantic vectors are available in **S** for both words. Figure 1 shows that there is a nonlinear relation between the correlation of words' semantic vectors in **S** and the ratings. The plot shows as well that for low correlations, the variability in the MEN ratings is larger. We fitted a Gaussian location scale additive model to this data set, summarized in Table 2, which supported r as a predictor for both mean MEN rating and the variability in the MEN ratings.

To put this performance in perspective, we collected the latent semantic analysis (LSA) similarity scores for the MEN word pairs using the website at `http://lsa.colorado.edu/`. The Spearman correlation for LSA scores and MEN ratings was 0.697, and that for r was 0.704 (both $p < 0.0001$). Thus, our semantic vectors perform on a par with those of LSA, a well-established older technique that still enjoys wide use in psycholinguistics. Undoubtedly, optimized techniques from computational linguistics such as `word2vec` will outperform our model. The important point here is that even with training on full sentences rather than using small windows and even when including function

Table 1: Linear mixed model fit to paired associate learning scores with the correlation r of the row vectors in **S** of the paired words as predictor. Treatment coding was used for factors. For the youngest age group, r is not predictive, but all other age groups show increasingly large slopes compared to the slope of the youngest age group. The range of r is [−4.88, −2.53]; hence larger values for the coefficients of r and its interactions imply worse performance on the PAL task.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept	3.6220	0.6822	5.3096	< 0.0001
r	0.3058	0.1672	1.8293	0.0691
age=39	0.2297	0.1869	1.2292	0.2207
age=49	0.4665	0.1869	2.4964	0.0135
age=59	0.6078	0.1869	3.2528	0.0014
age=69	0.8029	0.1869	4.2970	< 0.0001
sex=male	-0.1074	0.0230	-4.6638	< 0.0001
r:age=39	0.1167	0.0458	2.5490	0.0117
r:age=49	0.2090	0.0458	4.5640	< 0.0001
r:age=59	0.2463	0.0458	5.3787	< 0.0001
r:age=69	0.3239	0.0458	7.0735	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
random intercepts word pair	17.8607	18.0000	128.2283	< 0.0001

Table 2: Summary of a Gaussian location scale additive model fitted to the similarity ratings in the MEN dataset, with as predictor the correlation r of word's semantic vectors in the **S** matrix. TPRS: thin plate regression spline. b: the minimum standard deviation for the logb link function. Location: parameter estimating the mean; scale: parameter estimating the variance through the logb link function ($\eta = \log(\sigma - b)$).

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept [location]	25.2990	0.1799	140.6127	< 0.0001
Intercept [scale]	2.1297	0.0149	143.0299	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
TPRS r [location]	8.3749	8.8869	2766.2399	< 0.0001
TPRS r [scale]	5.9333	7.0476	87.5384	< 0.0001

words, performance of our linear network with linguistically motivated lexomes is sufficiently high to serve as a basis for further analysis.

3.2.3. Correlational Structure of Inflectional and Derivational Vectors. Now that we have established that the semantic vector space **S**, obtained with perhaps the simplest possible error-driven learning algorithm discriminating between outcomes given multiple cues (1), indeed captures semantic similarity, we next consider how the semantic vectors of inflectional and derivational lexomes cluster in this space. Figure 2 presents a heatmap for the correlation matrix of the functional lexomes that we implemented as described in Section 3.1.

Nominal and verbal inflectional lexomes, as well as adverbial ly, cluster in the lower left corner and tend to be either not correlated with derivational lexomes (indicated by light yellow) or to be negatively correlated (more reddish colors). Some inflectional lexomes, however, are interspersed with derivational lexomes. For instance, comparative, superlative, and perfective form a small subcluster together within the group of derivational lexomes.

The derivational lexomes show for a majority of pairs small positive correlations,and stronger correlations in the case of the verb-forming derivational lexomes mis, over, under, and undo, which among themselves show thev strongest positive correlations of all. The inflectional lexomes creating abstract nouns, ion, ence, ity, and ness also form a subcluster. In other words, Figure 2 shows that there is some structure to the distribution of inflectional and derivational lexomes in the semantic vector space.

Derived words (but not inflected words) have their own content lexomes and hence their own row vectors in **S**. Do the semantic vectors of these words cluster according to the their formatives? To address this question, we extracted the row vectors from **S** for a total of 3500 derived words for 31 different formatives, resulting in a 3500 × 23,562 matrix D sliced out of **S**. To this matrix, we added the semantic vectors for the 31 formatives. We then used linear discriminant analysis (LDA) to predict a lexome's formative from its semantic vector, using the `lda` function from the **MASS** package [82] (for LDA to work, we had to remove columns from **D** with very small variance; as a consequence, the dimension of the matrix that was the input to LDA was 3531 × 814). LDA accuracy for this classification task with 31 possible outcomes was 0.72. All 31 derivational lexomes were correctly classified.

When the formatives are randomly permuted, thus breaking the relation between the semantic vectors and their formatives, LDA accuracy was on average 0.528 (range across

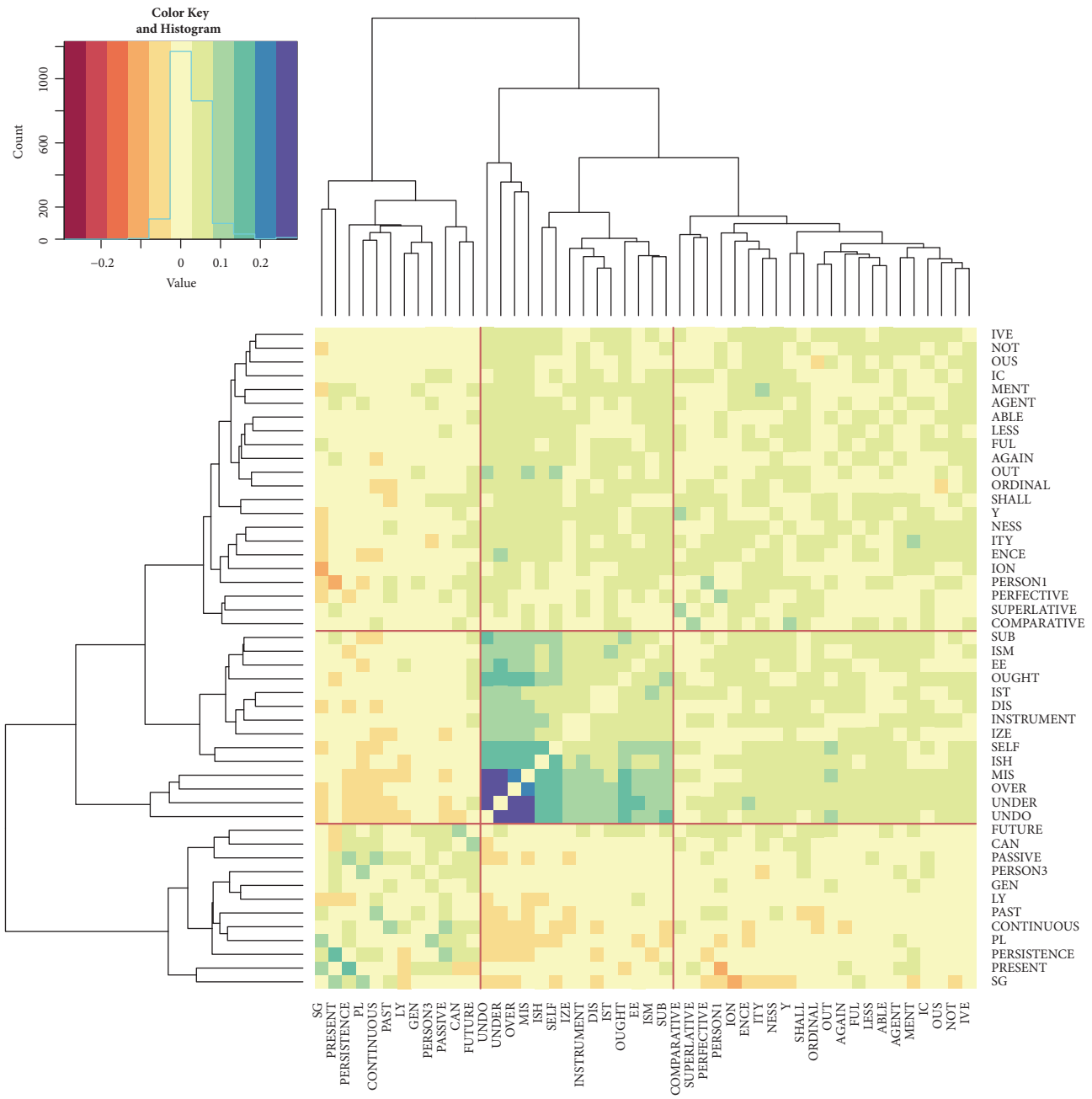


FIGURE 2: Heatmap for the correlation matrix of the row vectors in **S** of derivational and inflectional function lexomes (individual words will become legible by zooming in on the figure at maximal magnification.).

10 permutations 0.519–0.537). From this, we conclude that derived words show more clustering in the semantic space of **S** than can be expected under randomness.

How are the semantic vectors of derivational lexomes positioned with respect to the clusters of their derived words? To address this question, we constructed heatmaps for the correlation matrices of the pertinent semantic vectors. An example of such a heatmap is presented for NESS in Figure 3. Apart from the existence of clusters within the cluster of NESS content lexomes, it is striking that the NESS lexome itself is found at the very left edge of the dendrogram and at the very left column and bottom row of the heatmap. The color coding indicates that surprisingly the NESS derivational lexome is negatively correlated with almost all content lexomes that have *ness* as formative. Thus, the semantic vector of NESS is not a prototype at the center of its cloud of exemplars, but an *antiprototype*. This vector is close to the cloud of semantic vectors, but it is outside its periphery. This pattern is not specific to NESS but is found for the other derivational lexomes as well. It is intrinsic to our model.

The reason for this is straightforward. During learning, although for a derived word's lexome *i* and its derivational

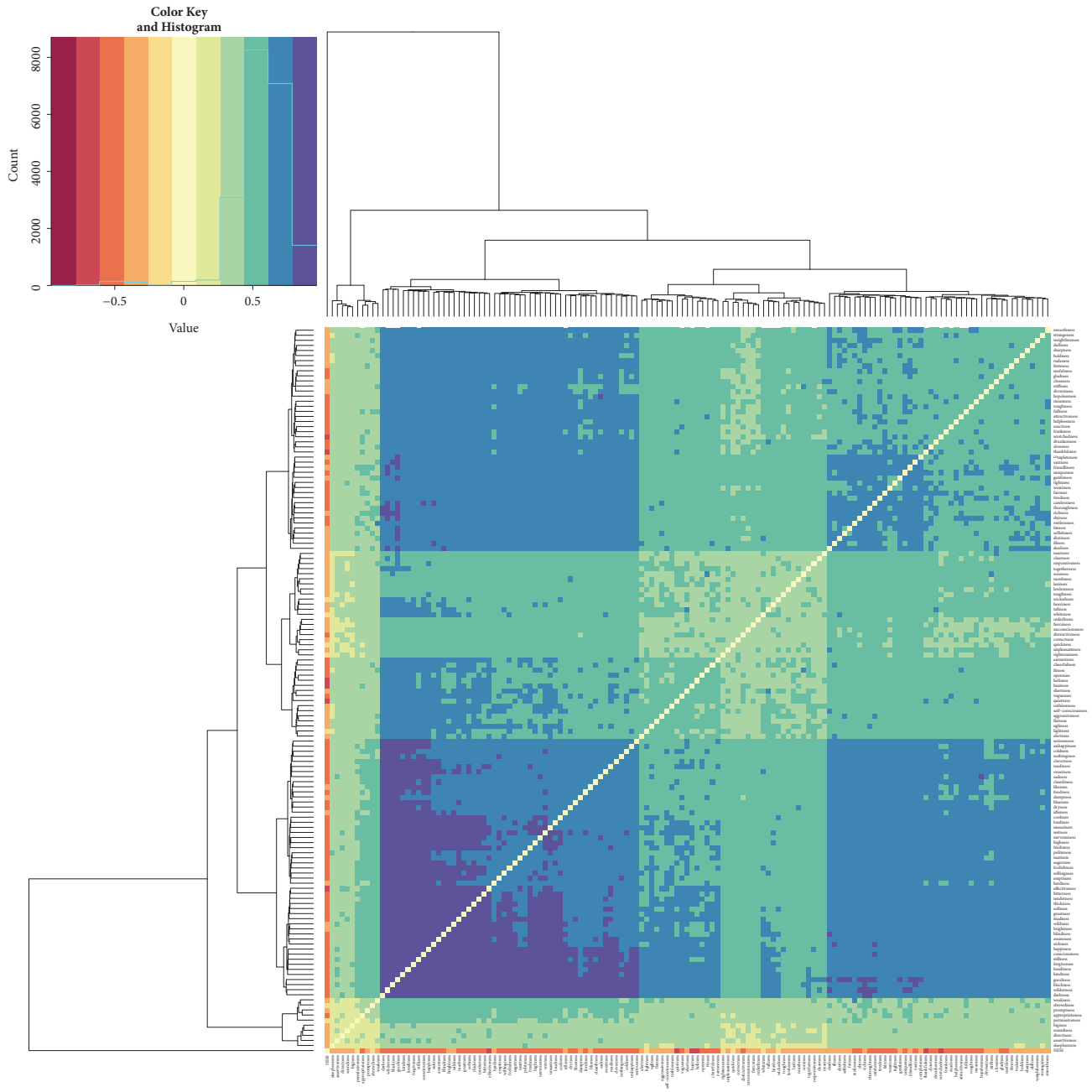


Figure 3: Heatmap for the correlation matrix of lexomes for content words with ness, as well as the derivational lexome of ness itself. This lexome is found at the very left edge of the dendrograms, and is negatively correlated with almost all content lexomes (Individual words will become legible by zooming in on the figure at maximal magnification.).

lexome ness are co-present cues, the derivational lexome occurs in many other words j, and each time another word j is encountered, weights are reduced from ness to i. As this happens for all content lexomes, the derivational lexome is, during learning, slowly but steadily discriminated away from its content lexomes. We shall see that this is an important property for our model to capture morphological productivity for comprehension and speech production.

When the additive model of Mitchell and Lapata [67] is used to construct a semantic vector for ness; i.e., when the average vector is computed for the vectors obtained by subtracting the vector of the derived word from that of the base word, the result is a vector that is embedded inside the cluster of derived vectors, and hence inherits semantic idiosyncracies from all these derived words.

3.2.4. Semantic Plausibility and Transparency Ratings for Derived Words. In order to obtain further evidence for the validity of inflectional and derivational lexomes, we re-analyzed the semantic plausibility judgements for word pairs

consisting of a base and a novel derived form (e.g., *accent*, *accentable*) reported by Marelli and Baroni [60] and available at http://clic.cimec.unitn.it/composes/FRACSS. The number of pairs available to us for analysis, 236, was restricted compared to the original 2,559 pairs due to the constraints that the base words had to occur in the TASA corpus. Furthermore, we also explored the role of words' emotional valence, arousal, and dominance, as available in Warriner et al. [83], which restricted the number of items even further. The reason for doing so is that human judgements, such as those of age of acquisition, may reflect dimensions of emotion [59].

A measure derived from the semantic vectors of the derivational lexomes, activation diversity, and the L1-norm of the semantic vector (the sum of the absolute values of the vector elements, i.e., its city-block distance) turned out to be predictive for these plausibility ratings, as documented in Table 3 and the left panel of Figure 4. Activation diversity is a measure of lexicality [58]. In an auditory word identification task, for instance, speech that gives rise to a low activation diversity elicits fast rejections, whereas speech that generates high activation diversity elicits higher acceptance rates but at the cost of longer response times [18].

Activation diversity interacted with word length. A greater word length had a strong positive effect on rated plausibility, but this effect progressively weakened as activation diversity increases. In turn, activation diversity had a positive effect on plausibility for shorter words, and a negative effect for longer words. Apparently, the evidence for lexicality that comes with higher L1-norms contributes positively when there is little evidence coming from word length. As word length increases and the relative contribution of the formative in the word decreases, the greater uncertainty that comes with higher lexicality (a greater L1-norm implies more strong links to many other lexomes) has a detrimental effect on the ratings. Higher arousal scores also contributed to higher perceived plausibility. Valence and dominance were not predictive and were therefore left out of the specification of the model reported here (word frequency was not included as predictor because all derived words are neologisms with zero frequency; addition of base frequency did not improve model fit, $p > 0.91$.).

The degree to which a complex word is semantically transparent with respect to its base word is of both practical and theoretical interest. An evaluation of the semantic transparency of complex words using transformation matrices is developed in Marelli and Baroni [60]. Within the present framework, semantic transparency can be examined straightforwardly by comparing the correlations of (i) the semantic vectors of the base word to which the semantic vector of the affix is added, with (ii) the semantic vector of the derived word itself. The more distant the two vectors are, the more negative their correlation should be. This is exactly what we find. For NESS, for instance, the six most negative correlations are *business* ($r = -0.66$), *effectiveness* ($r = -0.51$), *awareness* ($r = -0.50$), *loneliness* ($r = -0.45$), *sickness* ($r = -0.44$), and *consciousness* ($r = -0.43$). Although NESS can have an anaphoric function in discourse [84], words such as *business* and *consciousness* have a much deeper and richer semantics than just reference to a previously mentioned state of affairs. A simple comparison of the word's actual semantic vector in **S** (derived from the TASA corpus) and its semantic vector obtained by accumulating evidence over base and affix (i.e., summing the vectors of base and affix) brings this out straightforwardly.

However, evaluating correlations for transparency is prone to researcher bias. We therefore also investigated to what extent the semantic transparency ratings collected by Lazaridou et al. [68] can be predicted from the activation diversity of the semantic vector of the derivational lexome. The original dataset of Lazaridou and colleagues comprises 900 words, of which 330 meet the criterion of occurring more than 8 times in the TASA corpus and for which we have semantic vectors available. The summary of a Gaussian location-scale additive model is given in Table 4, and the right panel of Figure 4 visualizes the interaction of word length by activation diversity. Although modulated by word length, overall, transparency ratings decrease as activation diversity is increased. Apparently, the stronger the connections a derivational lexome has with other lexomes, the less clear it becomes what its actual semantic contribution to a novel derived word is. In other words, under semantic uncertainty, transparency ratings decrease.

4. Comprehension

Now that we have established that the present semantic vectors make sense, even though they are based on a small corpus, we next consider a comprehension model that has form vectors as input and semantic vectors as output (a package for R implementing the comprehension and production algorithms of linear discriminative learning is available at http://www.sfs.uni-tuebingen.de/~hbaayen/publications/WpmWithLdl_1.0.tar.gz. The package is described in Baayen et al. [85]). We begin with introducing the central concepts underlying mappings from form to meaning. We then discuss visual comprehension, and then turn to auditory comprehension.

4.1. Setting Up the Mapping. Let **C** denote the cue matrix, a matrix that specifies for each word (rows) the form cues of that word (columns). For a toy lexicon with the words *one, two,* and *three*, the **C** matrix is

$$\mathbf{C} = \begin{array}{c} \\ \text{one} \\ \text{two} \\ \text{three} \end{array} \begin{array}{c} \begin{array}{cccccccc} \#\text{wV} & \text{wVn} & \text{Vn\#} & \#\text{tu} & \text{tu\#} & \#\text{Tr} & \text{Tri} & \text{ri\#} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}, \tag{12}$$

where we use the DISC keyboard phonetic alphabet (the "DIstinct Single Character" representation with one character for each phoneme was introduced by CELEX [86]) for triphones; the # symbol denotes the word boundary. Suppose that the semantic vectors for these words are the row vectors

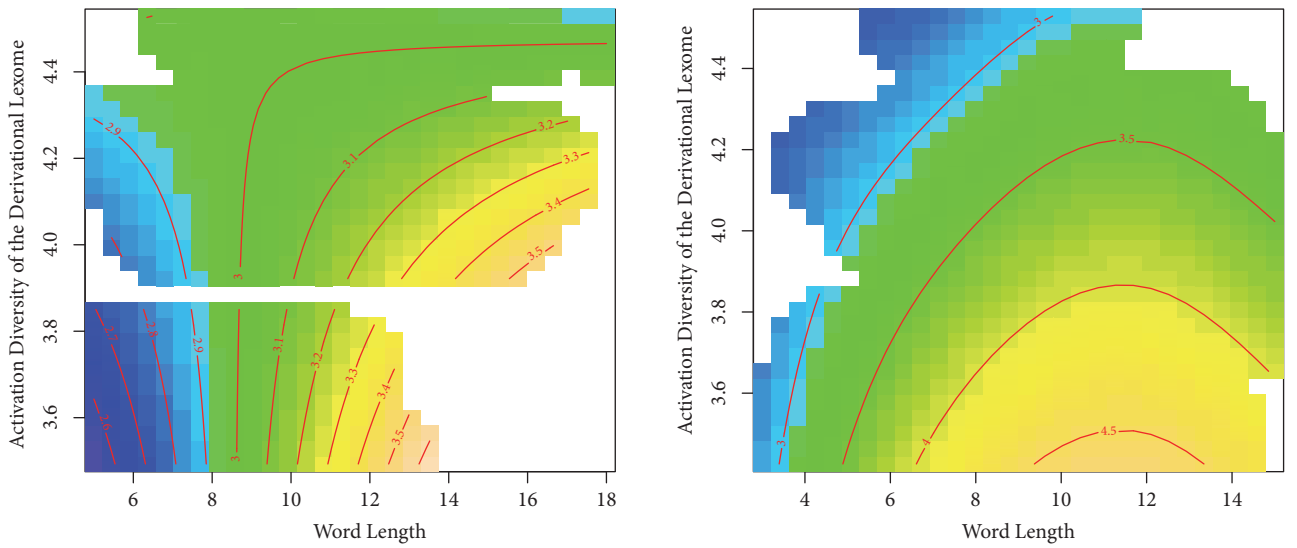


Figure 4: Interaction of word length by activation diversity in GAMs fitted to plausibility ratings for complex words (left) and to semantic transparency ratings (right).

Table 3: GAM fitted to plausibility ratings for derivational neologisms with the derivational lexomes able, again, agent, ist, less, and not; data from Marelli and Baroni [60].

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	-1.8072	2.7560	-0.6557	0.5127
Word Length	0.5901	0.2113	2.7931	0.0057
Activation Diversity	1.1221	0.6747	1.6632	0.0976
Arousal	0.3847	0.1521	2.5295	0.0121
Word Length × Activation Diversity	-0.1318	0.0500	-2.6369	0.0089
B. smooth terms	edf	Ref.df	F-value	p-value
Random Intercepts Affix	3.3610	4.0000	55.6723	< 0.0001

of the following matrix **S**:

$$\mathbf{S} = \begin{array}{c} \\ \text{one} \\ \text{two} \\ \text{three} \end{array} \begin{array}{c} \begin{array}{ccc} \text{one} & \text{two} & \text{three} \end{array} \\ \begin{pmatrix} 1.0 & 0.3 & 0.4 \\ 0.2 & 1.0 & 0.1 \\ 0.1 & 0.1 & 1.0 \end{pmatrix} \end{array}. \tag{13}$$

We are interested in a transformation matrix F such that

$$\mathbf{CF} = \mathbf{S}. \tag{14}$$

The transformation matrix is straightforward to obtain. Let $\mathbf{C}'$ denote the Moore-Penrose generalized inverse of **C**, available in R as the `ginv` function in the **MASS** package [82]. Then

$$\mathbf{F} = \mathbf{C}'\mathbf{S}. \tag{15}$$

For the present example,

$$\mathbf{F} = \begin{array}{c} \\ \text{\#wV} \\ \text{wVn} \\ \text{Vn\#} \\ \text{\#tu} \\ \text{tu\#} \\ \text{\#Tr} \\ \text{Tri} \\ \text{ri\#} \end{array} \begin{array}{c} \begin{array}{ccc} \text{one} & \text{two} & \text{three} \end{array} \\ \begin{pmatrix} 0.33 & 0.10 & 0.13 \\ 0.33 & 0.10 & 0.13 \\ 0.33 & 0.10 & 0.13 \\ 0.10 & 0.50 & 0.05 \\ 0.10 & 0.50 & 0.05 \\ 0.03 & 0.03 & 0.33 \\ 0.03 & 0.03 & 0.33 \\ 0.03 & 0.03 & 0.33 \end{pmatrix} \end{array}, \tag{16}$$

and for this simple example, **CF** is exactly equal to **S**.

In the remainder of this section, we investigate how well this very simple end-to-end model performs for visual word recognition as well as auditory word recognition. For visual word recognition, we use the semantic matrix **S** developed in Section 3, but we consider two different cue matrices **C**, one using letter trigrams (following [58]) and one using phone trigrams. A comparison of the performance of the two models

Table 4: Gaussian location-scale additive model fitted to the semantic transparency ratings for derived words. te: tensor product smooth and s: thin plate regression spline.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept [location]	5.5016	0.2564	21.4537	< 0.0001
intercept [scale]	-0.4135	0.0401	-10.3060	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(activation diversity, word length) [location]	5.3879	6.2897	23.7798	0.0008
s(derivational lexome) [location]	14.3180	15.0000	380.6254	< 0.0001
s(activation diversity) [scale]	2.1282	2.5993	28.4336	< 0.0001
s(word length) [scale]	1.7722	2.1403	132.7311	< 0.0001

sheds light on the role of words' "sound image" on word recognition in reading (cf. [87–89], for phonological effects in visual word Recognition). For auditory word recognition, we make use of the acoustic features developed in Arnold et al. [18].

Although the features that we selected as cues are based on domain knowledge, they do not have an ontological status in our framework, in contrast to units such as phonemes and morphemes in standard linguistic theories. In these theories, phonemes and morphemes are Russelian atomic units of formal phonological and syntactic calculi, and considerable research has been directed towards showing that these units are psychologically real. For instance, speech errors involving morphemes have been interpreted as clear evidence for the existence in the mind of morphemes [90]. Research has been directed at finding behavioral and neural correlates of phonemes and morphemes [6, 91, 92], ignoring fundamental criticisms of both the phoneme and the morpheme as theoretical constructs [25, 93, 94].

The features that we use for representing aspects of form are heuristic features that have been selected or developed primarily because they work well as discriminators. We will gladly exchange the present features for other features, if these other features can be shown to afford higher discriminative accuracy. Nevertheless, the features that we use are grounded in domain knowledge. For instance, the letter trigrams used for modeling reading (see, e.g.,[58]) are motivated by the finding in stylometry that letter trigrams are outstanding features for discriminating between authorial hands [95] (in what follows, we work with letter triplets and triphones, which are basically contextually enriched letter and phone units. For languages with strong phonotactic restrictions, such that the syllable inventory is quite small (e.g., Vietnamese), digraphs work appear to work better than trigraphs [96]. Baayen et al. [85] show that working with four-grams may enhance performance, and current work in progress on many other languages shows that for highly inflecting languages with long words, 4-grams may outperform 3-grams. For computational simplicity, we have not experimented with mixtures of units of different length, nor with algorithms with which such units might be determined.). Letter n-grams are also posited by Cohen and Dehaene [97] for the visual word form system, at the higher end of a hierarchy of neurons tuned to increasingly large visual features of words.

Triphones, the units that we use for representing the 'acoustic image' or 'auditory verbal imagery' of canonical word forms, have the same discriminative potential as letter triplets, but have as additional advantage that they do better justice, compared to phonemes, to phonetic contextual interdependencies, such as plosives being differentiated primarily by formant transitions in adjacent vowels. The acoustic features that we use for modeling auditory comprehension, to be introduced in more detail below, are motivated in part by the sensitivity of specific areas on the cochlea to different frequencies in the speech signal.

Evaluation of model predictions proceeds by comparing the predicted semantic vector $\hat{\mathbf{s}}$ obtained by multiplying an observed cue vector $\mathbf{c}$ with the transformation matrix $\mathbf{F}$ ($\hat{\mathbf{s}} = \mathbf{cF}$) with the corresponding target row vector $\mathbf{s}$ of $\mathbf{S}$. A word i is counted as correctly recognized when $\hat{\mathbf{s}}_i$ is most strongly correlated with the target semantic vector $\mathbf{s}_i$ of all target vectors $\mathbf{s}_j$ across all words j.

Inflected words do not have their own semantic vectors in $\mathbf{S}$. We therefore created semantic vectors for inflected words by adding the semantic vectors of stem and affix and added these as additional row vectors to $\mathbf{S}$ before calculating the transformation matrix $\mathbf{F}$.

We note here that there is only one transformation matrix, i.e., one discrimination network, that covers all affixes, inflectional and derivational, as well as derived and monolexomic (simple) words. This approach contrasts with that of Marelli and Baroni [60], who pair every affix with its own transformation matrix.

4.2. Visual Comprehension. For visual comprehension, we first consider a model straightforwardly mapping form vectors onto semantic vectors. We then expand the model with an indirect route first mapping form vectors onto the vectors for the acoustic image (derived from words' triphone representations) and then mapping the acoustic image vectors onto the semantic vectors.

The dataset on which we evaluated our models comprised 3987 monolexomic English words, 6595 inflected variants of monolexomic words, and 898 derived words with monolexomic base words, to a total of 11480 words. These counts follow from the simultaneous constraints of (i) a word appearing with sufficient frequency in TASA, (ii) the word being available in the British Lexicon Project (BLP [98]), (iii)

the word being available in the CELEX database, and the word being of the abovementioned morphological type. In the present study, we did not include inflected variants of derived words, nor did we include compounds.

4.2.1. The Direct Route Straight from Orthography to Semantics. To model single word reading as gauged by the visual lexical decision task, we used letter trigrams as cues. In our dataset, there were a total of 3465 unique trigrams, resulting in a 11480 × 3465 orthographic cue matrix **C**. The semantic vectors of the monolexomic and derived words were taken from the semantic weight matrix described in Section 3. For inflected words, semantic vectors were obtained by summation of the semantic vectors of base words and inflectional functions. For the resulting semantic matrix **S**, we retained the 5030 column vectors with the highest variances, setting the cutoff value for the minimal variance to 0.34×10^{-7}. From **S** and **C**, we derived the transformation matrix **F**, which we used to obtain estimates $\hat{\mathbf{s}}$ of the semantic vectors **s** in **S**.

For 59% of the words, $\hat{\mathbf{s}}$ had the highest correlation with the targeted semantic vector **s** (for the correctly predicted words, the mean correlation was 0.83, and the median was 0.86. With respect to the incorrectly predicted words, the mean and median correlation were both 0.48). The accuracy obtained with naive discriminative learning, using orthogonal lexomes as outcomes instead of semantic vectors, was 27%. Thus, as expected, performance of linear discrimination learning (LDL) is substantially better than that of naive discriminative learning (NDL).

To assess whether this model is productive, in the sense that it can make sense of novel complex words, we considered a separate set of inflected and derived words which were not included in the original dataset. For an unseen complex word, both base word and the inflectional or derivational function appeared in the training set, but not the complex word itself. The network **F** therefore was presented with a novel form vector, which it mapped onto a novel vector in the semantic space. Recognition was successful if this novel vector was more strongly correlated with the semantic vector obtained by summing the semantic vectors of base and inflectional or derivational function than with any other semantic vector.

Of 553 unseen inflected words, 43% were recognized successfully. The semantic vectors predicted from their trigrams by the network **F** were overall well correlated in the mean with the targeted semantic vectors of the novel inflected words (obtained by summing the semantic vectors of their base and inflectional function): $\overline{r} = 0.67, p < 0.0001$. The predicted semantic vectors also correlated well with the semantic vectors of the inflectional functions ($\overline{r} = 0.61, p < 0.0001$). For the base words, the mean correlation dropped to $\overline{r} = 0.28$ ($p < 0.0001$).

For unseen derived words (514 in total), we also calculated the predicted semantic vectors from their trigram vectors. The resulting semantic vectors $\hat{\mathbf{s}}$ had moderate positive correlations with the semantic vectors of their base words ($\overline{r} = 0.40, p < 0.0001$), but negative correlations with the semantic vectors of their derivational functions ($\overline{r} = -0.13, p < 0.0001$). They did not correlate with the semantic vectors obtained by summation of the semantic vectors of base and affix ($\overline{r} = 0.01, p = 0.56$).

The reduced correlations for the derived words as compared to those for the inflected words likely reflects to some extent that the model was trained on many more inflected words (in all, 6595) than derived words (898), whereas the number of different inflectional functions (7) was much reduced compared to the number of derivational functions (24). However, the negative correlations of the derivational semantic vectors with the semantic vectors of their derivational functions fit well with the observation in Section 3.2.3 that derivational functions are antiprototypes that enter into negative correlations with the semantic vectors of the corresponding derived words. We suspect that the absence of a correlation of the predicted vectors with the semantic vectors obtained by integrating over the vectors of base and derivational function is due to the semantic idiosyncracies that are typical for derived words. For instance, *austerity* can denote harsh discipline, or simplicity, or a policy of deficit cutting, or harshness to the taste. And a *worker* is not just someone who happens to work, but someone earning wages, or a nonreproductive bee or wasp, or a thread in a computer program. Thus, even within the usages of one and the same derived word, there is a lot of semantic heterogeneity that stands in stark contrast to the straightforward and uniform interpretation of inflected words.

4.2.2. The Indirect Route from Orthography via Phonology to Semantics. Although reading starts with orthographic input, it has been shown that phonology actually plays a role during the reading process as well. For instance, developmental studies indicate that children's reading development can be predicted by their phonological abilities [87, 99]. Evidence of the influence of phonology on adults' reading has also been reported [89, 100–105]. As pointed out by Perrone-Bertolotti et al. [106], silent reading often involves an imagery speech component: we hear our own "inner voice" while reading. Since written words produce a vivid auditory experience almost effortlessly, they argue that auditory verbal imagery should be part of any neurocognitive model of reading. Interestingly, in a study using intracranial EEG recordings with four epileptic neurosurgical patients, they observed that silent reading elicits auditory processing in the absence of any auditory stimulation, suggesting that auditory images are spontaneously evoked during reading (see also [107]).

To explore the role of auditory images in silent reading, we operationalized sound images by means of phone trigrams (in our model, phone trigrams are independently motivated as intermediate targets of speech production, see Section 5 for further discussion). We therefore ran the model again with phone trigrams instead of letter triplets. The semantic matrix **S** was exactly the same as above. However, a new transformation matrix **F** was obtained for mapping a 11480 × 5929 cue matrix **C** of phone trigrams onto **S**. To our surprise, the triphone model outperformed the trigram model substantially. Overall accuracy rose to 78%, an improvement of almost 20%.

In order to integrate this finding into our model, we implemented an additional network **K** that maps orthographic vectors (coding the presence or absence of trigrams) onto phonological vectors (indicating the presence or absence of triphones). The result is a dual route model for (silent) reading, with a first route utilizing a network mapping straight from orthography to semantics, and a second, indirect, route utilizing two networks, one mapping from trigrams to triphones, and the other subsequently mapping from triphones to semantics.

The network **K**, which maps trigram vectors onto triphone vectors, provides good support for the relevant triphones, but does not provide guidance as to their ordering. Using a graph-based algorithm detailed in the Appendix (see also Section 5.2), we found that for 92% of the words the correct sequence of triphones jointly spanning the word's sequence of letters received maximal support.

We then constructed a matrix $\widehat{\mathbf{T}}$ with the triphone vectors predicted from the trigrams by network **K**, and defined a new network **H** mapping these predicted triphone vectors onto the semantic vectors **S**. The mean correlation for the correctly recognized words was 0.90, and the median correlation was 0.94. For words that were not recognized correctly, the mean and median correlation were 0.45 and 0.43 respectively. We now have, for each word, two estimated semantic vectors, a vector $\widehat{\mathbf{s}}_1$ obtained with network **F** of the first route, and a vector $\widehat{\mathbf{s}}_2$ obtained with network **H** from the auditory targets that themselves were predicted by the orthographic cues. The mean of the correlations of these pairs of vectors was $\overline{r} = 0.73$ ($p < 0.0001$).

To assess to what extent the networks that we defined are informative for actual visual word recognition, we made use of the reaction times in the visual lexical decision task available in the BLP. All 11,480 words in the current dataset are also included in BLP. We derived three measures from each of the two networks.

The first measure is the sum of the L1-norms of $\widehat{\mathbf{s}}_1$ and $\widehat{\mathbf{s}}_2$, to which we will refer as a word's *total activation diversity*. The total activation diversity is an estimate of the support for a word's lexicality as provided by the two routes. The second measure is the correlation of $\widehat{\mathbf{s}}_1$ and $\widehat{\mathbf{s}}_2$, to which we will refer as a word's *route congruency*. The third measure is the L1-norm of a lexome's column vector in **S**, to which we refer as its *prior*. This last measure has previously been observed to be a strong predictor for lexical decision latencies [59]. A word's prior is an estimate of a word's entrenchment and prior availability.

We fitted a generalized additive model to the inverse-transformed response latencies in the BLP with these three measures as key covariates of interest, including as control predictors word length and word type (derived, inflected, and monolexomic, with derived as reference level). As shown in Table 5, inflected words were responded to more slowly than derived words, and the same holds for monolexomic words, albeit to a lesser extent. Response latencies also increased with length. Total activation diversity revealed a U-shaped effect (Figure 5, left panel). For all but the lowest total activation diversities, we find that response times increase as total activation diversity increases. This result is consistent with previous findings for auditory word recognition [18]. As expected given the results of Baayen et al. [59], the prior was a strong predictor, with larger priors affording shorter response times. There was a small effect of route congruency, which emerged in a nonlinear interaction with the prior, such that for large priors, a greater route congruency afforded further reduction in response time (Figure 5, right panel). Apparently, when the two routes converge on the same semantic vector, uncertainty is reduced and a faster response can be initiated.

For comparison, the dual-route model was implemented with NDL as well. Three measures were derived from the networks and used to predict the same response latencies. These measures include activation, activation diversity, and prior, all of which have been reported to be reliable predictors for RT in visual lexical decision [54, 58, 59]. However, since here we are dealing with two routes, the three measures can be independently derived from both routes. We therefore summed up the measures of the two routes, obtaining three measures: total activation, total activation diversity, and total prior. With word type and word length as control factors, Table 6 shows that these measures participated in a three-way interaction, presented in Figure 6. Total activation showed a U-shaped effect on RT that is increasingly attenuated as total activation diversity is increased (left panel). Total activation also interacted with the total prior (right panel). For medium range values of total activation, RTs increased with total activation, and as expected decreased with the total prior. The center panel shows that RTs decrease with total prior for most of the range of total activation diversity, and that the effect of total activation diversity changes sign going from low to high values of the total prior.

Model comparison revealed that the generalized additive model with LDL measures (Table 5) provides a substantially improved fit to the data compared to the GAM using NDL measures (Table 6), with a difference of no less than 651.01 AIC units. At the same time, the GAM based on LDL is less complex.

Given the substantially better predictability of LDL measures on human behavioral data, one remaining question is whether it is really the case that two routes are involved in silent reading. After all, the superior accuracy of the second route might be an artefact of a simple machine learning technique performing better for triphones than for trigrams. This question can be addressed by examining whether the fit of the GAM summarized in Table 5 improves or worsens depending on whether the activation diversity of the first or the second route is taken out of commission.

When the GAM is provided access to just the activation diversity of $\widehat{\mathbf{s}}_2$, the AIC increased by 100.59 units. However, when the model is based only on the activation diversity of $\widehat{\mathbf{s}}_1$, the model fit increased by no less than 147.90 AIC units. From this, we conclude that, at least for the visual lexical decision latencies in the BLP, the second route, first mapping trigrams to triphones, and then mapping triphones onto semantic vectors, plays the more important role.

The superiority of the second route may in part be due to the number of triphone features being larger

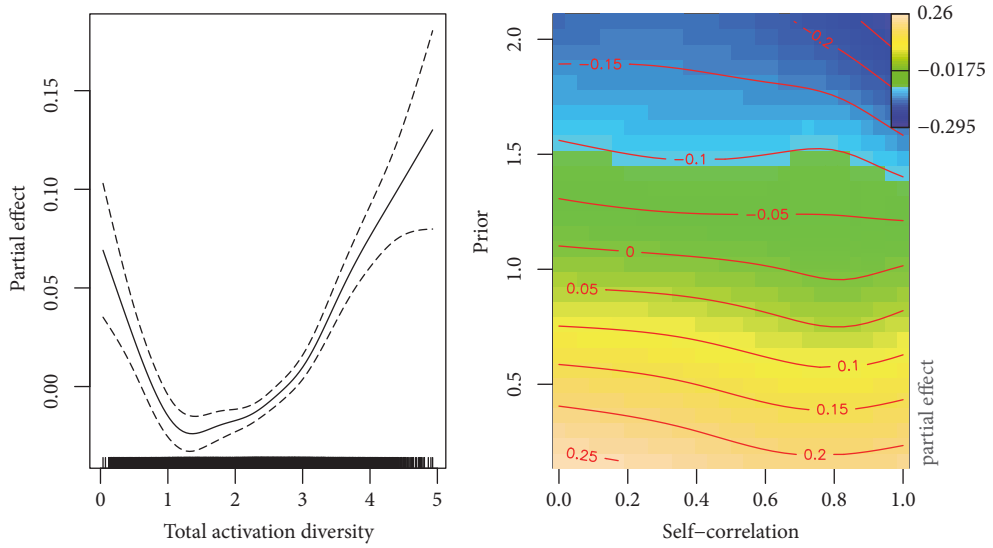


Figure 5: The partial effects of total activation diversity (left) and the interaction of route congruency and prior (right) on RT in the British Lexicon Project.

Table 5: Summary of a generalized additive model fitted to response latencies in visual lexical decision using measures based on LDL. s: thin plate regression spline smooth; te: tensor product smooth.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept	-1.7774	0.0087	-205.414	<.0001
word type:Inflected	0.1110	0.0059	18.742	<.0001
word type:Monomorphemic	0.0233	0.0054	4.322	<.0001
word length	0.0117	0.0011	10.981	<.0001
B. smooth terms	edf	Ref.df	F	p-value
s(total activation diversity)	6.002	7.222	23.6	<.0001
te(route congruency, prior)	14.673	17.850	213.7	<.0001

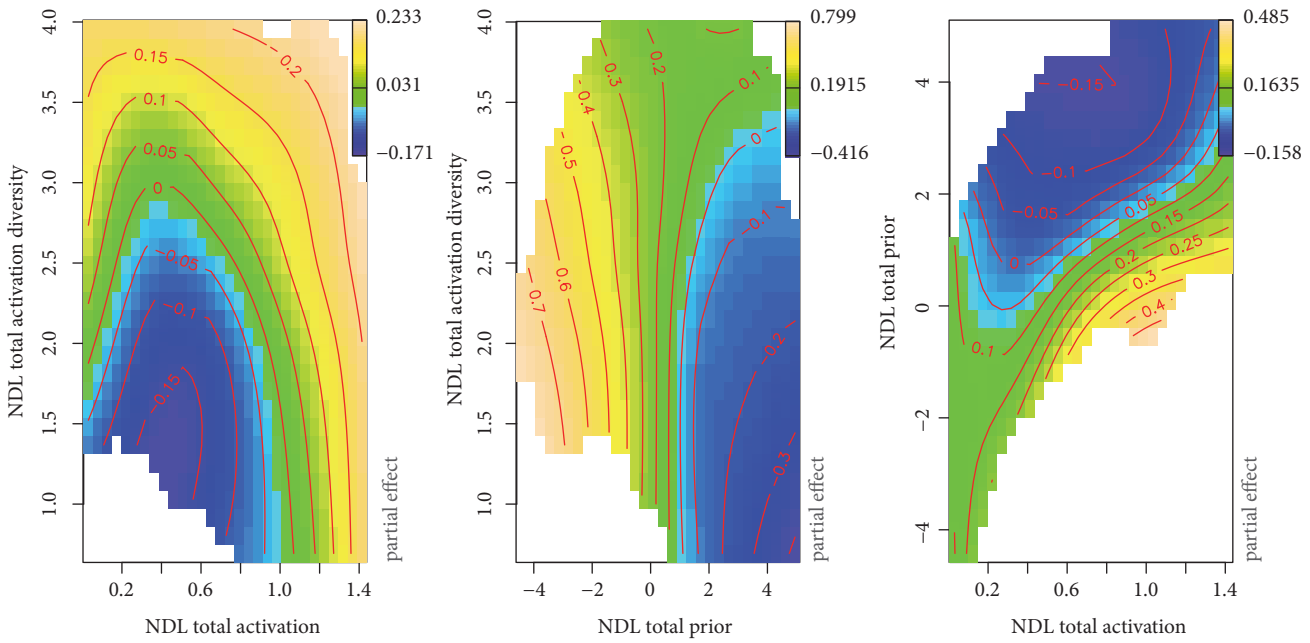


Figure 6: The interaction of total activation and total activation diversity (left), of total prior by total activation diversity (center), and of total activation by total prior (right) on RT in the British Lexicon Project, based on NDL.

Table 6: Summary of a generalized additive model fitted to response latencies in visual lexical decision using measures from NDL. te: tensor product smooth.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept	-1.6934	0.0086	-195.933	<.0001
word type:Inflected	0.0286	0.0052	5.486	<.0001
word type:Monomorphemic	0.0042	0.0055	0.770	= .4
word length	0.0066	0.0011	5.893	<.0001
B. smooth terms	edf	Ref.df	F	p-value
te(activation, activation diversity, prior)	41.34	49.72	108.4	<.0001

than the number of trigram features (3465 trigrams versus 5929 triphones). More features, which mathematically amount to more predictors, enable more precise mappings. Furthermore, the heavy use made in English of letter sequences such as *ough* (with 10 different pronunciations, `https://www.dictionary.com/e/s/ough`) reduces semantic discriminability compared to the corresponding spoken forms. It is noteworthy, however, that the benefits of the triphone-to-semantics mapping are possible only thanks to the high accuracy with which orthographic trigram vectors are mapped onto phonological triphone vectors (92%).

It is unlikely that the 'phonological route' is always dominant in silent reading. Especially in fast 'diagonal' reading, the 'direct route' may be more dominant. There is remarkable, although for the present authors, unexpected, convergence with the dual route model for reading aloud of Coltheart et al. [108] and Coltheart [109]. However, while the text-to-phonology route of their model has as primary function to explain why nonwords can be pronounced, our results show that both routes can actually be active when silently reading real words. A fundamental difference is, however, that in our model, words' semantics play a central role.

4.3. Auditory Comprehension. For the modeling of reading, we made use of letter trigrams as cues. These cues abstract away from the actual visual patterns that fall on the retina, patterns that are already transformed at the retina before being sent to the visual cortex. Our hypothesis is that letter trigrams represent those high-level cells or cell assemblies in the visual system that are critical for reading, and we therefore leave the modeling, possibly with deep learning networks of how patterns on the retina are transformed into letter trigrams for further research.

As a consequence of the high level of abstraction of the trigrams, a word form is represented by a unique vector specifying which of a fixed set of letter trigrams is present in the word. Although one might consider modeling auditory comprehension with phone triplets (triphones), replacing the letter trigrams of visual comprehension, such an approach would not do justice to the enormous variability of actual speech. Whereas the modeling of reading printed words can depart from the assumption that the pixels of a word's letters on a computer screen are in a fixed configuration, independently of where the word is shown on the screen, the speech signal of the same word type varies from token to token, as illustrated in Figure 7 for the English word *crisis*. A survey of the Buckeye corpus [110] of spontaneous conversations recorded at Columbus, Ohio [111], indicates that around 5% of the words are spoken with one syllable missing, and that a little over 20% of words have at least one phone missing.

It is widely believed that the phoneme, as an abstract unit of sound, is essential for coming to grips with the huge variability that characterizes the speech signal [112–114]. However, the phoneme as theoretical linguistic construct is deeply problematic [93], and for many spoken forms, canonical phonemes do not do justice to the phonetics of the actual sounds [115]. Furthermore, if words are defined as sequences of phones, the problem arises what representations to posit for words with two or more reduced variants. Adding entries for reduced forms to the lexicon turns out not to afford better overal recognition [116]. Although exemplar models have been put forward to overcome this problem [117], we take a different approach here and, following Arnold et al. [18], lay out a discriminative approach to auditory comprehension.

The cues that we make use of to represent the acoustic signal are the Frequency Band Summary Features (FBSFs) introduced by Arnold et al. [18] as input cues. FBSFs summarize the information present in the spectrogram of a speech signal. The algorithm that derives FBSFs first chunks the input at the minima of the Hilbert amplitude envelope of the signal's oscillogram (see the upper panel of Figure 8). For each chunk, the algorithm distinguishes 21 frequency bands in the MEL scaled spectrum, and intensities are discretized into 5 levels (lower panel in Figure 8) for small intervals of time. For each chunk, and for each frequency band in these chunks, a discrete feature is derived that specifies chunk number, frequency band number, and a description of the temporal variation in the band bringing together minimum, maximum, median, initial, and final intensity values. The 21 frequency bands are inspired by the 21 receptive areas on the cochlear membrane that are sensitive to different ranges of frequencies [118]. Thus, a given FBSF is a proxy for cell assemblies in the auditory cortex that respond to a particular pattern of changes over time in spectral intensity. The `AcousticNDLCodeR` package [119] for R [120] was employed to extract the FBSFs from the audio files.

We tested LDL on 20 hours of speech sampled from the audio files of the UCLA LIBRARY BROADCAST NEWSSCAPE data, a vast repository of multimodal TV news broadcasts,

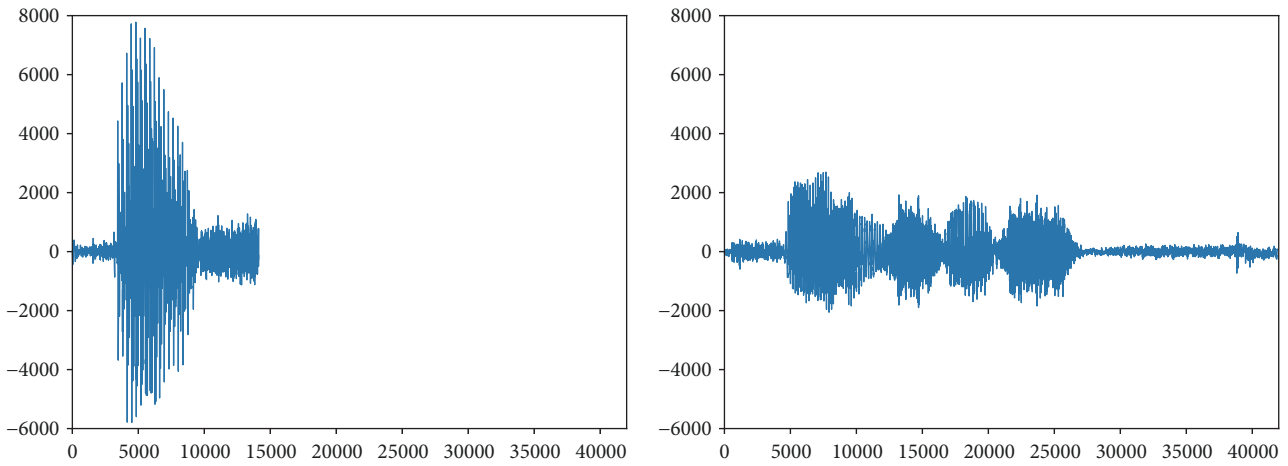


FIGURE 7: Oscillogram for two different realizations of the word *crisis* with different degrees of reduction in the NewsScape archive: [kraɪz](left) and [k^hraɪsɪs](right).

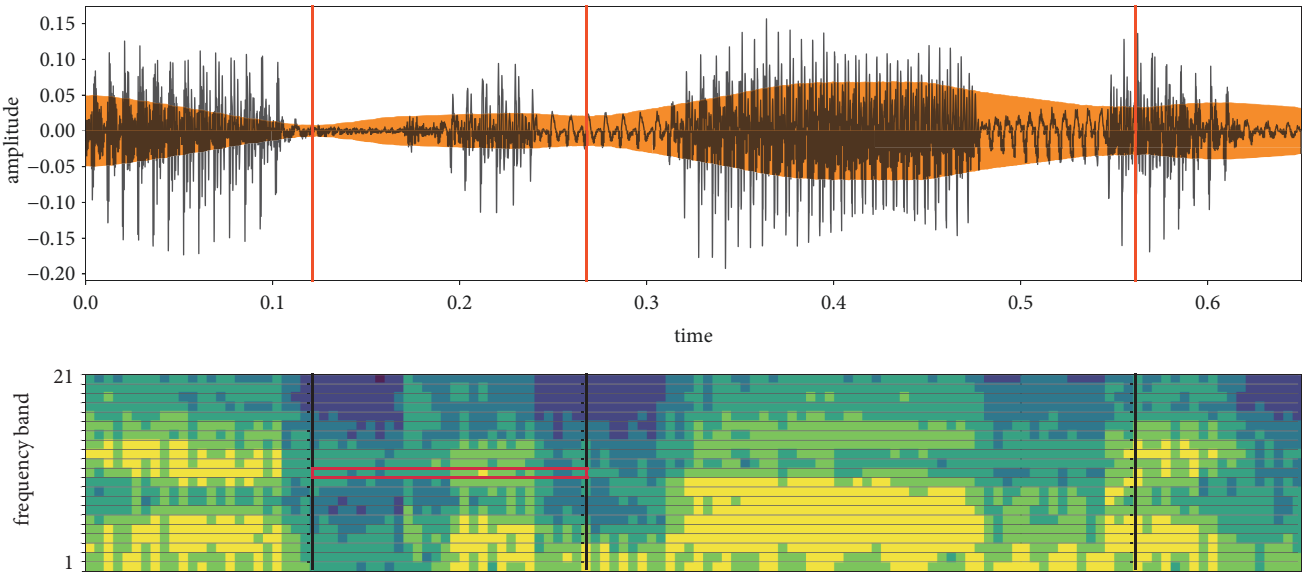


FIGURE 8: Oscillogram of the speech signal for a realization of the word *economic* with Hilbert envelope (in orange) outlining the signal is shown on the top panel. The lower panel depicts the discretized MEL scaled spectrum of the signal. The vertical bars are the boundary points that partition the signal into 4 chunks. For one chunk, horizontal lines (in red) highlight one of the frequency bands for which the FBSFs provide summaries of the variation over time in that frequency band. The FBSF for the highlighted band is "band11-start3-median3-min2-max5-end3-part2."

provided to us by the Distributed Little Red Hen Lab. The audio files of this resource were automatically classified as *clean* for relatively clean parts where there is speech without background noise or music and *noisy* for speech snippets where background noise or music is present. Here, we report results for 20 hours of clean speech, to a total of 131,673 word tokens (representing 4779 word types) with in all 40,639 distinct FBSFs.

The FBSFs for the word tokens are brought together in a matrix $\mathbf{C}_a$, with dimensions 131,673 audio tokens × 40,639 FBSFs. The targeted semantic vectors are taken from the **S** matrix, which is expanded to a matrix with 131,673 rows, one for each audio token, and 4,609 columns, the dimension of the semantic vectors. Although the transformation matrix **F** could be obtained by calculating $\mathbf{C}'\mathbf{S}$, the calculation of $\mathbf{C}'$ is numerically expensive. To reduce computational costs, we calculated **F** as follows: (the transpose of a square matrix **X**, denoted by $\mathbf{X}^T$ is obtained by replacing the upper triangle of the matrix by the lower triangle, and vice versa. Thus, $\left(\begin{smallmatrix}3 & 8\\ 7 & 2\end{smallmatrix}\right)^T = \left(\begin{smallmatrix}3 & 7\\ 8 & 2\end{smallmatrix}\right)$. For a nonsquare matrix, the transpose is obtained by switching rows and columns. Thus, a 4 × 2 matrix becomes a 2 × 4 matrix when transposed):

$$\mathbf{CF} = \mathbf{S}$$

$$\mathbf{C}^T\mathbf{CF} = \mathbf{C}^T\mathbf{S}$$

$$\left(\mathbf{C}^T\mathbf{C}\right)^{-1}\left(\mathbf{C}^T\mathbf{C}\right)\mathbf{F} = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{S}$$
$$\mathbf{F} = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{S}. \tag{17}$$

In this way, matrix inversion is required for a much smaller matrix $\mathbf{C}^T\mathbf{C}$, which is a square matrix of size 40,639 × 40,639.

To evaluate model performance, we compared the estimated semantic vectors with the targeted semantic vectors using the Pearson correlation. We therefore calculated the 131,673 × 131,673 correlation matrix for all possible pairs of estimated and targeted semantic vectors. Precision was calculated by comparing predicted vectors with the gold standard provided by the targeted vectors. Recognition was defined to be successful if the correlation of the predicted vector with the targeted gold vector was the highest of all the pairwise correlations of this predicted vector with any of the other gold semantic vectors. Precision, defined as the proportion of correct recognitions divided by the total number of audio tokens, was at 33.61%. For the correctly identified words, the mean correlation was 0.72, and for the incorrectly identified words, it was 0.55. To place this in perspective, a naive discrimination model with discrete lexomes as output performed at 12%, and a deep convolution network, Mozilla DeepSpeech (`https://github.com/mozilla/DeepSpeech`, based on Hannun et al. [9]), performed at 6%. The low performance of Mozilla DeepSpeech is due primarily due to its dependence on a language model. When presented with utterances instead of single words, it performs remarkably well.

4.4. Discussion. A problem with naive discriminative learning that has been puzzling for a long time is that measures based on NDL performed well as predictors of processing times [54, 58], whereas accuracy for lexical decisions was low. An accuracy of 27% for a model performing an 11,480-classification task is perhaps reasonable, but the lack of precision is unsatisfactory when the goal is to model human visual lexicality decisions. By moving from NDL to LDL, model accuracy is substantially improved (to 59%). At the same time, predictions for reaction times improved considerably as well. As lexicality decisions do not require word identification, further improvement in predicting decision behavior is expected to be possible by considering not only whether the predicted semantic is closest to the targeted vector but also measures such as how densely the space around the predicted semantic vector is populated.

Accuracy for auditory comprehension is lower, for the data we considered above at around 33%. Interestingly, a series of studies indicates that recognizing isolated words taken out of running speech is a nontrivial task also for human listeners [121–123]. Correct identification by native speakers of 1000 randomly sampled word tokens from a German corpus of spontaneous conversational speech ranged between 21% and 44% [18]. For both human listeners and automatic speech recognition systems, recognition improves considerably when words are presented in their natural context. Given that LDL with FBSFs performs very well on isolated word recognition, it seems worth investigating further whether the present approach can be developed into a fully-fledged model of auditory comprehension that can take full utterances as input. For a blueprint of how we plan to implement such a model, see Baayen et al. [124].

5. Speech Production

This section examines whether we can predict words' forms from the semantic vectors of **S**. If this is possible for the present dataset with reasonable accuracy, we have a proof of concept that discriminative morphology is feasible not only for comprehension, but also for speech production. The first subsection introduces the computational implementation. The next subsection reports on the model's performance, which is evaluated first for monomorphemic words, then for inflected words, and finally for derived words. Section 5.3 provides further evidence for the production network by showing that as the support from the semantics for the triphones becomes weaker, the amount of time required for articulating the corresponding segments increases.

5.1. Computational Implementation. For a production model, some representational format is required for the output that in turn drives articulation. In what follows, we make use of triphones as output features. Triphones capture part of the contextual dependencies that characterize speech and that render problematic the phoneme as elementary unit of a phonological calculus [93]. Triphones are in many ways not ideal, in that they inherit the limitations that come with discrete units. Other output features, structured along the lines of gestural phonology [125] or time series of movements of key articulators registered with electromagnetic articulography or ultrasound are on our list for further exploration. For now, we use triphones as a convenience construct, and we will show that given the support from the semantics for the triphones, the sequence of phones can be reconstructed with high accuracy. As some models of speech production assemble articulatory targets from phone segments (e.g., [126]), the present implementation can be seen as a front-end for this type of model.

Before putting the model to the test, we first clarify the way the model works by means of our toy lexicon with the words *one, two,* and *three*. Above, we introduced the semantic matrix **S** (13), which we repeat here for convenience,

$$\mathbf{S} = \begin{array}{c} \\ \text{one} \\ \text{two} \\ \text{three} \end{array} \begin{array}{c} \begin{array}{ccc} \text{one} & \text{two} & \text{three} \end{array} \\ \begin{pmatrix} 1.0 & 0.3 & 0.4 \\ 0.2 & 1.0 & 0.1 \\ 0.1 & 0.1 & 1.0 \end{pmatrix} \end{array} \tag{18}$$

as well as an **C** indicator matrix specifying which triphones occur in which words (12). As in what follows this matrix specifies the triphones targeted for production; we henceforth refer to this matrix as the **T** matrix.

$$\mathbf{T} = \begin{array}{c} \\ \text{one} \\ \text{two} \\ \text{three} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{\#wV} & \text{wVn} & \text{Vn\#} & \text{\#tu} & \text{tu\#} & \text{\#Tr} & \text{Tri} & \text{ri\#} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}. \tag{19}$$

For production, our interest is in the matrix **G** that transforms the row vectors of **S** into the row vectors of **T**; i.e., we need to solve

$$\mathbf{SG} = \mathbf{T}. \tag{20}$$

Given **G**, we can predict for any semantic vector **s** the vector of triphones $\hat{\mathbf{t}}$ that quantifies the support for the triphones provided by **s**, simply by multiplying **s** with **G**.

$$\hat{\mathbf{t}} = \mathbf{sG}. \tag{21}$$

As before, the transformation matrix **G** is straightforward to estimate. Let $\mathbf{S}'$ denote the Moore-Penrose generalized inverse of **S**. Since

$$\mathbf{S}'\mathbf{SG} = \mathbf{S}'\mathbf{T} \tag{22}$$

we have

$$\mathbf{G} = \mathbf{S}'\mathbf{T}. \tag{23}$$

Given **G**, we can predict the triphone matrix $\hat{\mathbf{T}}$ from the semantic matrix **S**:

$$\mathbf{SG} = \hat{\mathbf{T}}. \tag{24}$$

For the present example, the inverse of **S** and $\mathbf{S}'$ is

$$\mathbf{S}' = \begin{array}{c} \\ \text{one} \\ \text{two} \\ \text{three} \end{array} \begin{array}{c} \begin{array}{ccc} \text{one} & \text{two} & \text{three} \end{array} \\ \begin{pmatrix} 1.10 & -0.29 & -0.41 \\ -0.21 & 1.07 & -0.02 \\ -0.09 & -0.08 & 1.04 \end{pmatrix} \end{array} \tag{25}$$

and the transformation matrix **G** is

$$\mathbf{G} = \begin{array}{c} \\ \text{one} \\ \text{two} \\ \text{three} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{\#wV} & \text{wVn} & \text{Vn\#} & \text{\#tu} & \text{tu\#} & \text{\#Tr} & \text{Tri} & \text{ri\#} \end{array} \\ \begin{pmatrix} 1.10 & 1.10 & 1.10 & -0.29 & -0.29 & -0.41 & -0.41 & -0.41 \\ -0.21 & -0.21 & -0.21 & 1.07 & 1.07 & -0.02 & -0.02 & -0.02 \\ -0.09 & -0.09 & -0.09 & -0.08 & -0.08 & 1.04 & 1.04 & 1.04 \end{pmatrix} \end{array}. \tag{26}$$

For this simple example, $\hat{\mathbf{T}}$ is virtually identical to **T**. For realistic data, $\hat{\mathbf{T}}$ will not be identical to **T** but will be an approximation of it that is optimal in the least squares sense. The triphones with the strongest support are expected to be the most likely to be the triphones defining a word's form.

We made use of the **S** matrix, which we derived from the TASA corpus as described in Section 3. The majority of columns of the 23,561 × 23,561 matrix **S** show very small deviations from zero, and hence are uninformative. As before, we reduced the number of columns of **S** by removing columns with very low variance. Here, one option is to remove all columns with a variance below a preset threshold θ. However, to avoid adding a threshold as a free parameter, we set the number of columns retained to the number of different triphones in a given dataset, a number which is around n = 4500. In summary, **S** denotes a $w \times n$ matrix that specifies, for each of w words, an n-dimensional semantic vector.

5.2. Model Performance

5.2.1. Performance on Monolexomic Words. We first examined model performance on a dataset comprising monolexomic words that did not carry any inflectional exponents. This dataset of 3987 words comprised three irregular comparatives (*elder, less,* and *more*), two irregular superlatives (*least, most*), 28 irregular past tense forms, and one irregular past participle (*smelt*). For this set of words, we constructed a 3987 × 4446 matrix of semantic vectors $\mathbf{S}_m$ (a submatrix of the **S** matrix introduced above in Section 3) and a 3987 × 4446 triphone matrix $\mathbf{T}_m$ (a submatrix of **T**). The number of colums of $\mathbf{S}_m$ was set to the number of different triphones (the column dimension of $\mathbf{T}_m$). Those column vectors of $\mathbf{S}_m$ were retained that had the 4446 highest variances. We then estimated the transformation matrix **G** and used this matrix to predict the triphones that define words' forms.

We evaluated model performance in two ways. First, we inspected whether the triphones with maximal support were indeed the targeted triphones. This turned out to be the case for all words. Targeted triphones had an activation value close to one and nontargeted triphones an activation close to zero. As the triphones are not ordered, we also investigated whether the sequence of phones could be constructed correctly for these words. To this end, we set a threshold of 0.99, extracted all triphones with an activation exceeding this threshold, and used the `all_simple_paths` (a path is simple if the vertices it visits are not visited more than once.) function from the **igraph** package [127] to calculate all paths starting with any left-edge triphone in the set of extracted triphones. From the resulting set of paths, we selected the

longest path, which invariably was perfectly aligned with the sequence of triphones that defines words' forms.

We also evaluated model performance with a second, more general, heuristic algorithm that also makes use of the same algorithm from graph theory. Our algorithm sets up a graph with vertices collected from the triphones that are best supported by the relevant semantic vectors, and considers all paths it can find that lead from an initial triphone to a final triphone. This algorithm, which is presented in more detail in the appendix and which is essential for novel complex words, produced the correct form for 3982 out of 3987 words. It selected a shorter form for five words, *int* for *intent*, *lin* for *linnen*, *mis* for *mistress*, *oint* for *ointment*, and *pin* for *pippin*. The correct forms were also found but ranked second due to a simple length penalty that is implemented in the algorithm.

From these analyses, it is clear that mapping nearly 4000 semantic vectors on their corresponding triphone paths can be accomplished with very high accuracy for English monolexomic words. The question to be addressed next is how well this approach works for complex words. We first address inflected forms and limit ourselves here to the inflected variants of the present set of 3987 monolexomic words.

5.2.2. Performance on Inflected Words. Following the classic distinction between inflection and word formation, inflected words did not receive semantic vectors of their own. Nevertheless, we can create semantic vectors for inflected words by adding the semantic vector of an inflectional function to the semantic vector of its base. However, there are several ways in which the mapping from meaning to form for inflected words can be set up. To explain this, we need some further notation.

Let $\mathbf{S}_m$ and $\mathbf{T}_m$ denote the submatrices of $\mathbf{S}$ and $\mathbf{T}$ that contain the semantic and triphone vectors of monolexomic words. Assume that a subset of k of these monolexomic words is attested in the training corpus with inflectional function a. Let $\mathbf{T}_a$ denote the matrix with the triphone vectors of these inflected words, and let $\mathbf{S}_a$ denote the corresponding semantic vectors. To obtain $\mathbf{S}_a$, we take the pertinent submatrix $\mathbf{S}_{m_a}$ from $\mathbf{S}_m$ and add the semantic vector $\mathbf{s}_a$ of the affix:

$$\mathbf{S}_a = \mathbf{S}_{m_a} + \mathbf{i} \bigotimes \mathbf{s}_a. \tag{27}$$

Here, $\mathbf{i}$ is a unit vector of length k and $\bigotimes$ is the generalized Kronecker product, which in (27) stacks k copies of $\mathbf{s}_a$ row-wise. As a first step, we could define a separate mapping $\mathbf{G}_a$ for each inflectional function a,

$$\mathbf{S}_a\mathbf{G}_a = \mathbf{T}_a, \tag{28}$$

but in this set-up, learning of inflected words does not benefit from the knowledge of the base words. This can be remedied by a mapping for augmented matrices that contain the row vectors for both base words and inflected words:

$$\begin{bmatrix} \mathbf{S}_m \\ \mathbf{S}_a \end{bmatrix} \mathbf{G}_a = \begin{bmatrix} \mathbf{T}_m \\ \mathbf{T}_a \end{bmatrix}. \tag{29}$$

The dimension of $\mathbf{G}_a$ (length of semantic vector by length of triphone vector) remains the same, so this option is not more costly than the preceding one. Nevertheless, for each inflectional function, a separate large matrix is required. A much more parsimonious solution is to build augmented matrices for base words and all inflected words jointly:

$$\begin{bmatrix} \mathbf{S}_m \\ \mathbf{S}_{a_1} \\ \mathbf{S}_{a_2} \\ \vdots \\ \mathbf{S}_{a_n} \end{bmatrix} \mathbf{G} = \begin{bmatrix} \mathbf{T}_m \\ \mathbf{T}_{a_1} \\ \mathbf{T}_{a_2} \\ \vdots \\ \mathbf{T}_{a_n} \end{bmatrix} \tag{30}$$

The dimension of $\mathbf{G}$ is identical to that of $\mathbf{G}_a$, but now all inflectional functions are dealt with by a single mapping. In what follows, we report the results obtained with this mapping.

We selected 6595 inflected variants which met the criterion that the frequency of the corresponding inflectional function was at least 50. This resulted in a dataset with 91 comparatives, 97 superlatives, 2401 plurals, 1333 continuous forms (e.g., *walking*), 859 past tense forms and 1086 forms classed as perfective (past participles), as well as 728 third person verb forms (e.g., *walks*). Many forms can be analyzed as either past tenses or as past participles. We followed the analyses of the treetagger, which resulted in a dataset in which both inflectional functions are well-attested.

Following (30), we obtained an (augmented) 10582 × 5483 semantic matrix $\mathbf{S}$, whereas before we retained the 5483 columns with the highest column variance. The (augmented) triphone matrix $\mathbf{T}$ for this dataset had the same dimensions.

Inspection of the activations of the triphones revealed that targeted triphones had top activations for 85% of the monolexomic words and 86% of the inflected words. The proportion of words with at most one intruding triphone was 97% for both monolexomic and inflected words. The graph-based algorithm performed with an overall accuracy of 94%, accuracies broken down by morphology revealed an accuracy of 99% for the monolexomic words, and an accuracy of 92% for inflected words. One source of errors for the production algorithm is inconsistent coding in the CELEX database. For instance, the stem of *prosper* is coded as having a final schwa followed by r, but the inflected forms are coded without the r, creating a mismatch between a partially rhotic stem and completely non-rhotic inflected variants.

We next put model performance to a more stringent test by using 10-fold cross-validation for the inflected variants. For each fold, we trained on all stems and 90% of all inflected forms and then evaluated performance on the 10% of inflected forms that were not seen in training. In this way, we can ascertain the extent to which our production system (network plus graph-based algorithm for ordering triphones) is productive. We excluded from the cross-validation procedure irregular inflected forms, forms with CELEX phonological forms with inconsistent within-paradigm rhoticism, as well as forms the stem of which was not available in the training set. Thus,

cross-validation was carried out for a total of 6236 inflected forms.

As before, the semantic vectors for inflected words were obtained by addition of the corresponding content and inflectional semantic vectors. For each training set, we calculated the transformation matrix **G** from the **S** and **T** matrices of that training set. For an out-of-bag inflected form in the test set, we calculated its semantic vector and multiplied this vector with the transformation matrix (using (21)) to obtain the predicted triphone vector $\hat{\mathbf{t}}$.

The proportion of forms that were predicted correctly was 0.62. The proportion of forms ranked second was 0.25. Forms that were incorrectly ranked first typically were other inflected forms (including bare stems) that happened to receive stronger support than the targeted form. Such errors are not uncommon in spoken English. For instance, in the Buckeye corpus [110], *closest* is once reduced to *clos*. Furthermore, Dell [90] classified forms such as *concludement* for *conclusion*, and *he relax* for *he relaxes*, as (noncontextual) errors.

The algorithm failed to produce the targeted form for 3% of the cases. Examples of the forms produced instead are the past tense for *blazed* being realized with [zId] instead of [zd], the plural *mouths* being predicted as having [Ts] as coda rather than [Ds], and *finest* being reduced to *finst*. The voiceless production for *mouth* does not follow the dictionary norm, but is used as attested by on-line pronunciation dictionaries. Furthermore, the voicing alternation is partly unpredictable (see, e.g., [128], for final devoicing in Dutch), and hence model performance here is not unreasonable. We next consider model accuracy for derived words.

5.2.3. Performance on Derived Words. When building the vector space model, we distinguished between inflection and word formation. Inflected words did not receive their own semantic vectors. By contrast, each derived word was assigned its own lexome, together with a lexome for its derivational function. Thus, *happiness* was paired with two lexomes, HAPPINESS and NESS. Since the semantic matrix for our dataset already contains semantic vectors for derived words, we first investigated how well forms are predicted when derived words are assessed along with monolexomic words, without any further differentiation between the two. To this end, we constructed a semantic matrix **S** for 4885 words (rows) by 4993 (columns) and constructed the transformation matrix **G** from this matrix and the corresponding triphone matrix **T**. The predicted form vectors of $\hat{\mathbf{T}} = \mathbf{SG}$ supported the targeted triphones above all other triphones without exception. Furthermore, the graph-based algorithm correctly reconstructed 99% of all forms, with only 5 cases where it assigned the correct form second rank.

Next, we inspected how the algorithm performs when the semantic vectors of derived forms are obtained from the semantic vectors of their base words and those of their derivational lexomes, instead of using the semantic vectors of the derived words themselves. To allow subsequent evaluation by cross-validation, we selected those derived words that contained an affix that occured at least 30 times in our data set (AGAIN (38), AGENT (177), FUL (45), INSTRUMENT (82), LESS (54), LY (127), and NESS (57)), to a total of 770 complex words.

We first combined these derived words with the 3987 monolexomic words. For the resulting 4885 words, the **T** and **S** matrices were constructed, from which we derived the transformation matrix **G** and subsequently the matrix of predicted triphone strengths $\hat{\mathbf{T}}$. The proportion of words for which the targeted triphones were the best supported triphones was 0.96, and the graph algorithm performed with an accuracy of 98.9%.

In order to assess the productivity of the system, we evaluated performance on derived words by means of 10-fold cross-validation. For each fold, we made sure that each affix was present proportionally to its frequency in the overall dataset and that a derived word's base word was included in the training set.

We first examined performance when the transformation matrix is estimated from a semantic matrix that contains the semantic vectors of the derived words themselves, whereas semantic vectors for unseen derived words are obtained by summation of the semantic vectors of base and affix. It turns out that this set-up results in a total failure. In the present framework, the semantic vectors of derived words are too idiosyncratic and too finely tuned to their own collocational preferences. They are too scattered in semantic space to support a transformation matrix that supports the triphones of both stem and affix.

We then used exactly the same cross-validation procedure as outlined above for inflected words, constructing semantic vectors for derived words from the semantic vectors of base and affix, and calculating the transformation matrix **G** from these summed vectors. For unseen derived words, **G** was used to transform the semantic vectors for novel derived words (obtained by summing the vectors of base and affix) into triphone space.

For 75% of the derived words, the graph algorithm reconstructed the targeted triphones. For 14% of the derived nouns, the targeted form was not retrieved. These include cases such as *resound*, which the model produced with [s] instead of the (unexpected) [z], *sewer*, which the model produced with `@R` instead of only `R`, and *tumbler*, where the model used syllabic [l] instead of nonsyllabic [l] given in the CELEX target form.

5.3. Weak Links in the Triphone Graph and Delays in Speech Production. An important property of the model is that the support for trigrams changes where a word's graph branches out for different morphological variants. This is illustrated in Figure 9 for the inflected form *blending*. Support for the stem-final triphone `End` is still at 1, but then *blend* can end, or continue as *blends*, *blended*, or *blending*. The resulting uncertainty is reflected in the weights on the edges leaving `End`. In this example, the targeted *ing* form is driven by the inflectional semantic vector for CONTINUOUS, and hence the edge to `ndI` is best supported. For other forms, the edge weights will be different, and hence other paths will be better supported.

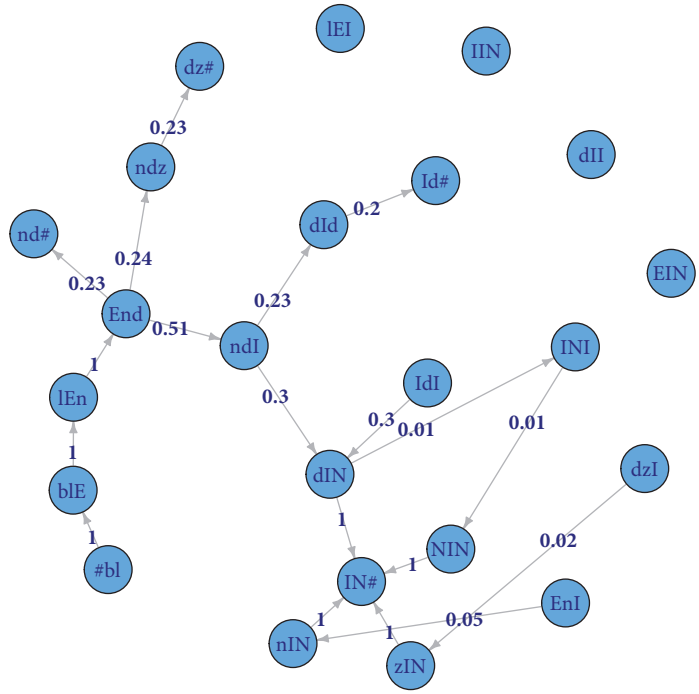


FIGURE 9: The directed graph for *blending*. Vertices represent triphones, including triphones such as IIN that were posited by the graph algorithm to bridge potentially relevant transitions that are not instantiated in the training data. Edges are labelled with their edge weights. The graph, the edge weights of which are specific to the lexomes BLEND and CONTINUOUS, incorporates not only the path for *blending*, but also the paths for *blend*, *blends*, and *blended*.

The uncertainty that arises at branching points in the triphone graph is of interest in the light of several experimental results. For instance, inter keystroke intervals in typing become longer at syllable and morph boundaries [129, 130]. Evidence from articulography suggests variability is greater at morph boundaries [131]. Longer keystroke execution times and greater articulatory variability are exactly what is expected under reduced edge support. We therefore examined whether the edge weights at the first branching point are predictive for lexical processing. To this end, we investigated the acoustic duration of the segment at the center of the first triphone with a reduced LDL edge weight (in the present example, d in ndI, henceforth 'branching segment') for those words in our study that are attested in the Buckeye corpus [110].

The dataset we extracted from the Buckeye corpus comprised 15105 tokens of a total of 1327 word types, collected from 40 speakers. For each of these words, we calculated the relative duration of the branching segment, calculated by dividing segment duration by word duration. For the purposes of statistical evaluation, the distribution of this relative duration was brought closer to normality by means of a logarithmic transformation. We fitted a generalized additive mixed model [132] to log relative duration with random intercepts for word and speaker, log LDL edge weight as predictor of interest and local speech rate (Phrase Rate), log neighborhood density (Ncount), and log word frequency as control variables. A summary of this model is presented in Table 7. As expected, an increase in LDL Edge Weight goes hand in hand with a reduction in the duration of the branching segment. In other words, when the edge weight is reduced, production is slowed.

The adverse effects of weaknesses in a word's path in the graph where the path branches is of interest against the discussion about the function of weak links in diphone transitions in the literature on comprehension [133–138]. For comprehension, it has been argued that bigram or diphone 'troughs,' i.e., low transitional probabilities in a chain of high transitional probabilities, provide points where sequences are segmented and parsed into their constituents. From the perspective of discrimination learning, however, low-probability transitions function in exactly the opposite way for comprehension [54, 124]. Naive discriminative learning also predicts that troughs should give rise to shorter processing times in comprehension, but not because morphological decomposition would proceed more effectively. Since high-frequency boundary bigrams and diphones are typically used word-internally across many words, they have a low cue validity for these words. Conversely, low-frequency boundary bigrams are much more typical for very specific base+affix combinations, and hence are better discriminative cues that afford enhanced activation of words' lexomes. This, in turn, gives rise to faster processing (see also Ramscar et al. [78] for the discriminative function of low-frequency bigrams in low-frequency monolexomic words).

There is remarkable convergence between the directed graphs for speech production such as illustrated in Figure 9 and computational models using temporal self-organizing

Table 7: Statistics for the partial effects of a generalized additive mixed model fitted to the relative duration of edge segments in the Buckeye corpus. The trend for log frequency is positive accelerating. TPRS: thin plate regression spline.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	-1.9165	0.0592	-32.3908	< 0.0001
Log LDL Edge Weight	-0.1358	0.0400	-3.3960	0.0007
Phrase Rate	0.0047	0.0021	2.2449	0.0248
Log Ncount	0.1114	0.0189	5.8837	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
TPRS log Frequency	1.9007	1.9050	7.5495	0.0004
random intercepts word	1149.8839	1323.0000	31.4919	< 0.0001
random intercepts speaker	30.2995	39.0000	34.5579	< 0.0001

maps (TSOMs, [139–141]). TSOMs also use error-driven learning, and in recent work attention is also drawn to weak edges in words' paths in the TSOM and the consequences thereof for lexical processing [142]. We are not using TSOMs, however, one reason being that in our experience they do not scale up well to realistically sized lexicons. A second reason is that we are pursuing the hypothesis that form serves meaning, and that self-organization of form by itself is not necessary. This hypothesis is likely too strong, especially as we do not provide a rationale for the trigram and triphone units that we use to represent aspects of form. It is worth noting that spatial organization of form features is not restricted to TSOMs. Although in our approach, the spatial organization of triphone features is left unspecified, such an organization can be easily enforced (see [85], for further details) by using an algorithm such as graphopt, which self-organizes the vertices of a graph into a two-dimensional plane. Figure 9 was obtained with this algorithm (Graphopt was developed for the layout of large graphs `http://www.schmuhl.org/graphopt/` and is implemented in the **igraph** package. Graphopt uses basic principles of physics to iteratively determine an optimal layout. Each node in the graph is given both mass and an electric charge, and edges between nodes are modeled as springs. This sets up a system in which there are attracting and repelling forces between the vertices of the graph, and this physical system is simulated until it reaches an equilibrium.).

5.4. Discussion. Our production network reconstructs known words' forms with a high accuracy, 99.9% for monolexomic words, 92% for inflected words, and 99% for derived words. For novel complex words, accuracy under 10-fold cross-validation was 62% for inflected words and 75% for derived words.

The drop in performance for novel forms is perhaps unsurprising, given that speakers understand many more words than they themselves produce, even though they hear novel forms on a fairly regular basis as they proceed through life [79, 143].

However, we also encountered several technical problems that are not due to the algorithm but to the representations that the algorithm has had to work with. First, it is surprising that accuracy is as high as it is given that the semantic vectors are constructed from a small corpus with a very simple discriminative algorithm. Second, we encountered inconsistencies in the phonological forms retrieved from the CELEX database, inconsistencies that in part are due to the use of discrete triphones. Third, many cases where the model predictions do not match the targeted triphone sequence, the targeted forms have a minor irregularity (e.g., *resound* with [z] instead of [s]). Fourth, several of the typical errors that the model makes are known kinds of speech errors or reduced forms that one might encounter in engaged conversational speech.

It is noteworthy that the model is almost completely data-driven. The triphones are derived from CELEX and other than some manual corrections for inconsistencies are derived automatically from words' phone sequences. The semantic vectors are based on the TASA corpus and were not in any way optimized for the production model. Given the triphone and semantic vectors, the transformation matrix is completely determined. No by-hand engineering of rules and exceptions is required, nor is it necessary to specify with hand-coded links what the first segment of a word is, what its second segment is, etc., as in the WEAVER model [37]. It is only in the heuristic graph algorithm that three thresholds are required, in order to avoid that graphs become too large to remain computationally tractable.

The speech production system that emerges from this approach comprises first of all an excellent memory for forms that have been encountered before. Importantly, this memory is not a static memory, but a dynamic one. It is not a repository of stored forms, but it reconstructs the forms from the semantics it is requested to encode. For regular unseen forms, however, a second network is required that projects a regularized semantic space (obtained by accumulation of the semantic vectors of content and inflectional or derivational functions) onto the triphone output space. Importantly, it appears that no separate transformations or rules are required for individual inflectional or derivational functions.

Jointly, the two networks, both of which can also be trained incrementally using the learning rule of Widrow-Hoff [44], define a dual route model; attempts to build a single integrated network were not successful. The semantic vectors of derived words are too idiosyncratic to allow generalization for novel forms. It is the property of the semantic vectors of derived lexomes described in Section 3.2.3, namely, that they are close to but not inside the cloud of their content

lexomes, which makes them productive. Novel forms do not partake in the idiosyncracies of lexicalized existing words but instead remain bound to base and derivational lexomes. It is exactly this property that makes them pronounceable. Once produced, a novel form will then gravitate as experience accumulates towards the cloud of semantic vectors of its morphological category, meanwhile also developing its own idiosyncracies in pronunciation, including sometimes highly reduced pronunciation variants (e.g., /tyk/ for Dutch *natuurlijk* ([natyrl@k]) [111, 144, 145]. It is perhaps possible to merge the two routes into one, but for this, much more fine-grained semantic vectors are required that incorporate information about discourse and the speaker's stance with respect to the adressee (see, e.g., [115], for detailed discussion of such factors).

6. Bringing in Time

Thus far, we have not considered the role of time. The audio signal comes in over time, longer words are typically read with more than one fixation, and likewise articulation is a temporal process. In this section, we briefly outline how time can be brought into the model. We do so by discussing the reading of proper names.

Proper names (morphologically compounds) pose a challenge to compositional theories, as the people referred to by names such as Richard Dawkins, Richard Nixon, Richard Thompson, and Sandra Thompson are not obviously semantic composites of each other. We therefore assign lexomes to names, irrespective of whether the individuals referred to are real or fictive, alive or dead. Furthermore, we assume that when the personal name and family name receive their own fixations, both names are understood as pertaining to the same named entity, which is therefore coupled with its own unique lexome. Thus, for the example sentences in Table 8, the letter trigrams of *John* are paired with the lexome JohnClark, and likewise the letter trigrams of *Clark* are paired with this lexome.

By way of illustration, we obtained the matrix of semantic vectors training on 100 randomly ordered tokens of the sentences of Table 8. We then constructed a trigram cue matrix $\mathbf{C}$ specifying, for each word (*John, Clark, wrote, ...*), which trigrams it contains. In parallel, a matrix $\mathbf{L}$ specifying for each word its corresponding lexomes (JohnClark, JohnClark, write, ...) was set up. We then calculated the matrix $\mathbf{F}$ by solving $\mathbf{CF} = \mathbf{L}$, and used $\mathbf{F}$ to calculate estimated (predicted) semantic vectors $\widehat{\mathbf{L}}$. Figure 10 presents the correlations of the estimated semantic vectors for the word forms *John*, *John Welsch*, *Janet*, and *Clark* with the targeted semantic vectors for the named entities JohnClark, JohnWelsch, JohnWiggam, AnneHastie, and JaneClark. For the sequentially read words *John* and *Welsch*, the semantic vector generated by *John* and that generated by *Welsch* were summed to obtain the integrated semantic vector for the composite name.

Figure 10, upper left panel, illustrates that upon reading the personal name *John*, there is considerable uncertainty about which named entity is at issue. When subsequently the family name is read (upper right panel), uncertainty is reduced and *John Welsch* now receives full support, whereas other named entities have correlations close to zero, or even negative correlations. As in the small world of the present toy example *Janet* is a unique personal name, there is no uncertainty about what named entity is at issue when *Janet* is read. For the family name *Clark* on its own, by contrast, *John Clark* and *Janet Clark* are both viable options.

For the present example, we assumed that every orthographic word received one fixation. However, depending on whether the eye lands far enough into the word, names such as *John Clark* and compounds such as *graph theory* can also be read with a single fixation. For single-fixation reading, learning events should comprise the joint trigrams of the two orthographic words, which are then simultaneously calibrated against the targeted lexome (JohnClark, graph-theory).

Obviously, the true complexities of reading, such as parafoveal preview, are not captured by this example. Nevertheless, as a rough first approximation, this example shows a way forward for integrating time into the discriminative lexicon.

7. General Discussion

We have presented a unified discrimination-driven model for visual and auditory word comprehension and word production: the discriminative lexicon. The layout of this model is visualized in Figure 11. Input (cochlea and retina) and output (typing and speaking) systems are presented in light gray; the representations of the model are shown in dark gray. For auditory comprehension, we make use of frequency band summary (FBS) features, low-level cues that we derive automatically from the speech signal. The FBS features serve as input to the auditory network $\mathbf{F}_a$ that maps vectors of FBS features onto the semantic vectors of $\mathbf{S}$. For reading, letter trigrams represent the visual input at a level of granularity that is optimal for a functional model of the lexicon (see [97], for a discussion of lower-level visual features). Trigrams are relatively high-level features compared to the FBS features. For features implementing visual input at a much lower level of visual granularity, using histograms of oriented gradients [146], see Linke et al. [15]. Letter trigrams are mapped by the visual network $\mathbf{F}_o$ onto $\mathbf{S}$.

For reading, we also implemented a network, here denoted by $\mathbf{K}_a$, that maps trigrams onto auditory targets (auditory verbal images), represented by triphones. These auditory triphones in turn are mapped by network $\mathbf{H}_a$ onto the semantic vectors $\mathbf{S}$. This network is motivated not only by our observation that for predicting visual lexical decision latencies, triphones outperform trigrams by a wide margin. Network $\mathbf{H}_a$ is also part of the control loop for speech production, see Hickok [147] for discussion of this control loop and the necessity of auditory targets in speech production. Furthermore, the acoustic durations with which English speakers realize the stem vowels of English verbs [148], as well as the duration of word final *s* in English [149],

Table 8: Example sentences for the reading of proper names. To keep the example simple, inflectional lexomes are not taken into account.

John Clark wrote great books about ants.
John Clark published a great book about dangerous ants.
John Welsch makes great photographs of airplanes.
John Welsch has a collection of great photographs of airplanes landing.
John Wiggam will build great harpsichords.
John Wiggam will be a great expert in tuning harpsichords.
Anne Hastie has taught statistics to great second year students.
Anne Hastie has taught probability to great first year students.
Janet Clark teaches graph theory to second year students.
JohnClark, write, great, book, about, ant
JohnClark, publish, a, great, book, about, dangerous, ant
JohnWelsch, make, great, photograph, of, airplane
JohnWelsch, have, a, collection, of, great, photograph, airplane, landing
JohnWiggam, future, build, great, harpsichord
JohnWiggam, future, be, a, great, expert, in, tuning, harpsichord
AnneHastie, have, teach, statistics, to, great, second, year, student
AnneHastie, have, teach, probability, to, great, first, year, student
JanetClark, teach, graphtheory, to, second, year, student

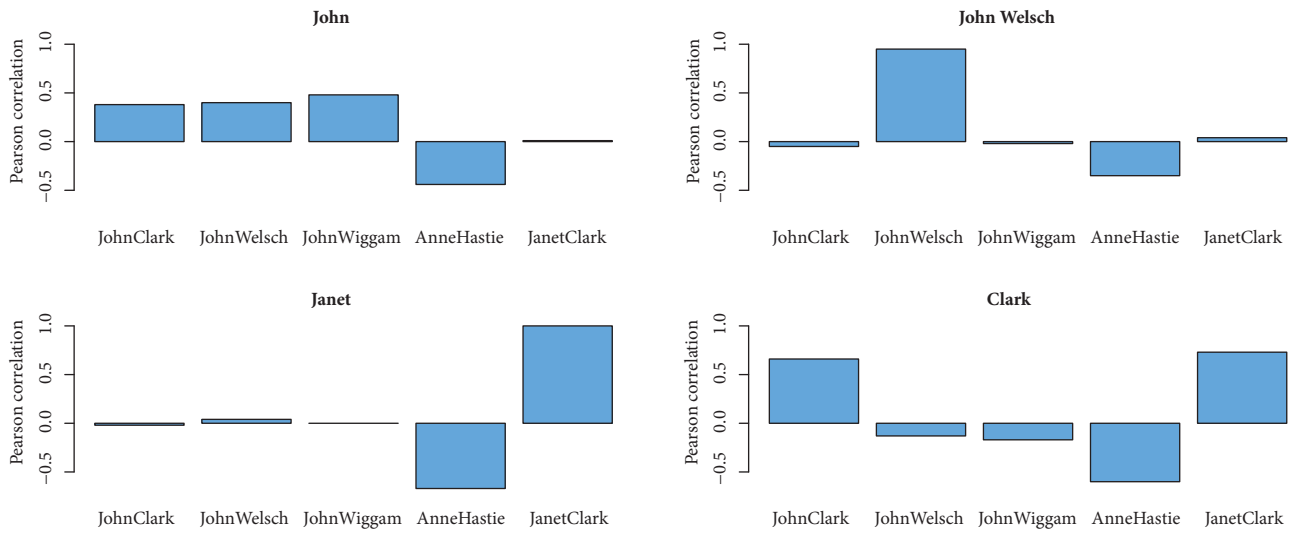


Figure 10: Correlations of estimated semantic vectors for *John*, *John Welsch*, *Janet*, and *Clark* with the targeted semantic vectors of JohnClark, JohnWelsch, JohnWiggam, AnneHastie, and JaneClark.

can be predicted from NDL networks using auditory targets to discriminate between lexomes [150].

Speech production is modeled by network $\mathbf{G}_a$, which maps semantic vectors onto auditory targets, which in turn serve as input for articulation. Although current models of speech production assemble articulatory targets from phonemes (see, e.g., [126, 147]), a possibility that is certainly compatible with our model when phones are recontextualized as triphones, we think that it is worthwhile investigating whether a network mapping semantic vectors onto vectors of articulatory parameters that in turn drive the articulators can be made to work, especially since many reduced forms are not straightforwardly derivable from the phone sequences of their 'canonical' forms [111, 144].

We have shown that the networks of our model have reasonable to high production and recognition accuracies, and also generate a wealth of well-supported predictions for lexical processing. Central to all networks is the hypothesis that the relation between form and meaning can be modeled discriminatively with large but otherwise surprisingly simple linear networks, the underlying mathematics of which (linear algebra) is well-understood. Given the present results, it is clear that the potential of linear algebra for understanding the lexicon and lexical processing has thus far been severely underestimated (the model as outlined in Figure 11 is a modular one, for reasons of computational convenience and interpretability. The different networks probably interact (see for some evidence [47]). One such interaction is explored

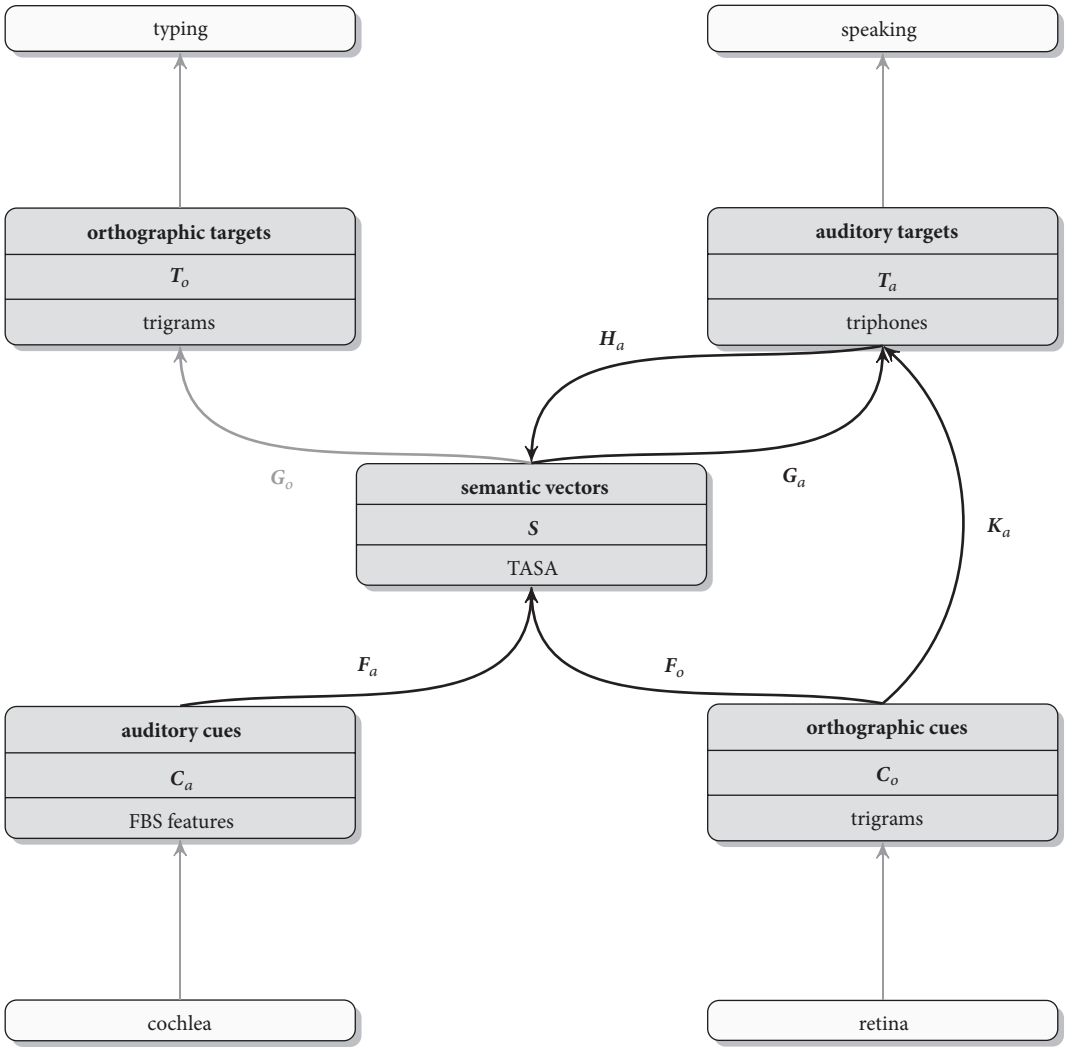


Figure 11: Overview of the discriminative lexicon. Input and output systems are presented in light gray, the representations of the model are shown in dark gray. Each of these representations is a matrix with its own set of features. Arcs are labeled with the discriminative networks (transformation matrices) that map matrices onto each other. The arc from semantic vectors to orthographic vectors is in gray, as this mapping is not explored in the present study.

in Baayen et al. [85]. In their production model for Latin inflection, feedback from the projected triphone paths back to the semantics (synthesis by analysis) is allowed, so that of otherwise equally well-supported paths, the path that best expresses the targeted semantics is selected. For information flows between systems in speech production, see Hickok [147].). In what follows, we touch upon a series of issues that are relevant for evaluating our model.

Incremental learning. In the present study, we estimated the weights on the connections of these networks with matrix operations, but, importantly, these weights can also be estimated incrementally, using the learning rule of Widrow and Hoff [44]; further improvements in accuracy are expected when using the Kalman filter [151]. As all matrices can be updated incrementally (and this holds as well for the matrix with semantic vectors, which are also time-variant) and in theory should be updated incrementally whenever information about the order of learning events is available, the present theory has potential for studying lexical acquisition and the continuing development of the lexicon over the lifetime [79].

Morphology without compositional operations. We have shown that our networks are predictive for a range of experimental findings. Important from the perspective of discriminative linguistics is that there are no compositional operations in the sense of Frege and Russell [1, 2, 152]. The work on compositionality in logic has deeply influenced formal linguistics (e.g., [3, 153]), and has led to the belief that the "architecture of the language faculty" is grounded in a homomorphism between a calculus (or algebra) of syntactic representations and a calculus (or algebra) based on semantic primitives. Within this tradition, compositionality arises when rules combining representations of form are matched with rules combining representations of meaning.

The approach of Marelli and Baroni [60], who derive the semantic vector of a complex word from the semantic vector of its base and a dedicated affix-specific mapping from the pertinent base words to the corresponding derived words, is in the spirit of compositionality, in that a change in form (through affixation) goes hand in hand with a change in meaning (modeled with a linear mapping in semantic space).

The perspective on morphological processing developed in the present study breaks with this tradition. Although semantic vectors for inflected words are obtained by summation of the semantic vectors of the base word and those of its inflectional functions (see [85], for modeling of the rich verbal paradigms of Latin), the form vectors corresponding to the resulting semantic vectors are not obtained by the summation of vectors for chunks of form, nor by any other operations on forms. Attempts to align chunks of words' form with their semantic structures — typically represented by graphs constructed over semantic primitives (see, e.g., [154]) — are destined to end in quagmires of arbitrary decisions about the number of inflectional classes required (Estonian, 30 or 300?), what the chunks should look like (does German have a morpheme *d* in its articles and demonstratives *der, die, das, diese, . . .*?), whether Semitic morphology requires operations on form on one or two tiers [155–157], and how many affixal position slots are required for Athebascan languages such as Navajo [158]. A language that is particularly painful for compositional theories is Yélî Dnye, an Indo-Pacific language spoken on Rossel island (in the south-east of Papua New Guinea). The language has a substantial inventory of over a 1000 function words used for verb agreement. Typically, these words are monosyllables that simultaneously specify values for negation, tense, aspect, person and number of subject, deixis, evidentiality, associated motion, and counterfactuality. Furthermore, verbs typically have no less than eight different stems, the use of which depends on particular constellations of inflectional functions [159, 160]. Yélî Dnye thus provides a striking counterexample to the hypothesis that a calculus of meaning would be paralleled by a calculus of form.

Whereas the theory of linear discriminative learning radically rejects the idea that morphological processing is grounded in compositionality as defined in mainstream formal linguistic theory, it does allow for the conjoint expression in form of lexical meanings and inflectional functions and for forms to map onto such conjoint semantic vectors. But just as multiple inflectional functions can be expressed jointly in a single word, multiple words can map onto nonconjoint semantic vectors, as was illustrated above for named entity recognition: *John Wiggam* can be understood to refer to a very specific person without this person being a compositional function of the proper name *John* and the surname *Wiggam*. In both these cases, there is no isomorphy between inflectional functions and lexical meanings on the one hand and chunks of form on the other hand.

Of course, conjoint vectors for lexical meanings and inflectional functions can be implemented and made to work only because a Latin verb form such as *sapīvissēmus* is understood, when building the semantic space **S**, to express lexomes for person (first), number (plural), voice (active), tense (pluperfect), mood (subjunctive), and lexical meaning (to know) (see [85], for computational modeling details). Although this could loosely be described as a decompositional approach to inflected words, we prefer not to use the term 'decompositional' as in formal linguistics this term, as outlined above, has a very different meaning. A more adequate terminology would be that the system is conjoint realizational for production and conjoint inferential for comprehension (we note here that we have focused on temporally conjoint realization and inference, leaving temporally disjoint realization and inference for sequences of words, and possibly compounds and fixed expression, for further research).

Naive versus linear discriminative learning. There is one important technical difference between the learning engine of the current study, LDL, and naive discriminative learning (NDL). Because NDL makes the simplifying assumption that outcomes are orthogonal, the weights from the cues to a given outcome are independent of the weights from the cues to the other outcomes. This independence assumption, which motivates the word *naive* in naive discriminative learning, makes it possible to very efficiently update network weights during incremental learning. In LDL, by contrast, this independence no longer holds: learning is no longer naive. We therefore refer to our networks not as NDL networks but simply as *linear* discriminative learning networks. Actual incremental updating of LDL networks is computationally more costly, but further numerical optimization is possible and implementations are underway.

Scaling. An important property of the present approach is that good results are obtained already for datasets of modest size. Our semantic matrix is derived from the TASA corpus, which with only 10 million words is dwarfed by the 2.8 billion words of the ukWaC corpus used by Marelli and Baroni [60], more words than any individual speaker can ever encounter in their lifetime. Similarly, Arnold et al. [18] report good results for word identification when models are trained on 20 hours of speech, which contrasts with the huge volumes of speech required for training deep learning networks for speech recognition. Although performance increases somewhat with more training data [161], it is clear that considerable headway can be made with relatively moderate volumes of speech. It thus becomes feasible to train models on, for instance, Australian English and New Zealand English, using already existing speech corpora, and to implement networks that make precise quantitative predictions for how understanding is mediated by listeners' expectations about their interlocutors (see, e.g., [162]).

In this study, we have obtained good results with a semantic matrix with a dimensionality of around 4000 lexomes. By itself, we find it surprising that a semantic vector space can be scaffolded with a mere 4000 words. But this finding may also shed light on the phenomenon of childhood amnesia [163, 164]. Bauer and Larkina [165] pointed out that young children have autobiographical memories that they may not remember a few years later. They argue that the rate of forgetting diminishes as we grow older and stabilizes in adulthood. Part of this process of forgetting may relate to the changes in the semantic matrix. Crucially, we expect the correlational structure of lexomes to change

considerably over time as experience accumulates. If this is indeed the case, the semantic vectors for the lexomes for young children are different from the corresponding lexomes for older children, which will again differ from those of adults. As a consequence, autobiographical memories that were anchored to the lexomes at a young age can no longer be accessed as the semantic vectors to which these memories were originally anchored no longer exist.

No examplars. Given how we estimate the transformation matrices with which we navigate between form and meaning, it might seem that our approach is in essence an implementation of an exemplar-based theory of the mental lexicon. After all, for instance, the $\mathbf{C}_a$ matrix specifies, for each auditory word form (exemplar), its pertinent frequency band summary features. However, the computational implementation of the present study should not be confused with the underlying algorithmic conceptualization. The underlying conceptualization is that learning proceeds incrementally, trial by trial, with the weights of transformation networks being updated with the learning rule of Widrow and Hoff [44]. In the mean, trial-by-trial updating with the Widrow-Hoff learning rule results in the same expected connection weights as obtained by the present estimates using the generalized inverse. The important conceptual advantage of incremental learning is, however, that there is no need to store exemplars. By way of example, for auditory comprehension, acoustic tokens are given with the input to the learning system. These tokens, the ephemeral result of interacting with the environment, leave their traces in the transformation matrix $\mathbf{F}_a$, but are not themselves stored. The same holds for speech production. We assume that speakers dynamically construct the semantic vectors from their past experiences, rather than retrieving these semantic vectors from some dictionary-like fixed representation familiar from current file systems. The dynamic nature of these semantic vectors emerged loud and clear from the modeling of novel complex words for speech production under cross-validation. In other words, our hypothesis is that all input and output vectors are ephemeral, in the sense that they represent temporary states of the cognitive system that are not themselves represented in that system but that leave their traces in the transformation matrices that jointly characterize and define the cognitive system that we approximate with the discriminative lexicon.

This dynamic perspective on the mental lexicon as consisting of several coupled networks that jointly bring the cognitive system into states that in turn feed into further networks (not modeled here) for sentence integration (comprehension) and articulation (production) raises the question about the status of units in this theory. Units play an important role when constructing numeric vectors for form and meaning. Units for letter trigrams and phone trigrams are, of course, context-enriched letter and phone units. Content lexomes as well as lexomes for derivational and inflectional functions are crutches for central semantic functions ranging from functions realizing relatively concrete onomasiological functions ('this is a dog, not a cat') to abstract situational functions allowing entities, events, and states to be specified on the dimensions of time, space, quantity, aspect, etc. However, crucial to the central thrust of the present study, there are no morphological symbols (units combining form and meaning) nor morphological form units such as stems and exponents of any kind in the model. Above, we have outlined the many linguistic considerations that have led us not to want to incorporate such units as part of our theory. Importantly, as there are no hidden layers in our model, it is not possible to seek for 'subsymbolic' reflexes of morphological units in patterns of activation over hidden layers. Here, we depart from earlier connectionist models, which rejected symbolic representations but retained the belief that there should be morphological representation, albeit subsymbolic ones. Seidenberg and Gonnerman [166], for instance, discuss morphological representations as being 'interlevel representations' that are emergent reflections of correlations among orthography, phonology, and semantics. This is not to say that the more agglutinative a language is, the more the present model will seem, to the outside observer, to be operating with morphemes. But the more a language tends towards fusional or polysynthetic morphology, the less strong this impression will be.

Model complexity. Next consider the question of how difficult lexical processing actually is. Sidestepping the complexities of the neural embedding of lexical processing [126, 147], here we narrow down this question to algorithmic complexity. State-of-the-art models in psychology [37, 114, 167] implement many layers of hierarchically organized units, and many hold it for an established fact that such units are in some sense psychologically real (see, e.g., [19, 91, 168]). However, empirical results can be equally well understood without requiring the theoretical construct of the morpheme (see, e.g., [47, 54, 58, 166, 169, 170]). The present study illustrates this point for 'boundary' effects in written and oral production. When paths in the triphone graph branch, uncertainty increases and processing slows down. Although observed 'boundary' effects are compatible with a post-Bloomfieldian lexicon, the very same effects arise in our model, even though morphemes do not figure in any way in our computations.

Importantly, the simplicity of the linear mappings in our model of the discriminative lexicon allows us to sidestep the problem that typically arise in computationally implemented full-scale hierarchical systems, namely, that errors arising at lower levels propagate to higher levels. The major steps forward made in recent end-to-end models in language engineering suggest that end-to-end cognitive models are also likely to perform much better. One might argue that the hidden layers of deep learning networks represent the traditional 'hidden layers' of phonemes, morphemes, and word forms mediating between form and meaning in linguistic models and offshoots thereof in psychology. However, the case of baboon lexical learning [12, 13] illustrates that this research strategy is not without risk, and that simpler discriminative networks can substantially outperform deep learning networks [15]. We grant, however, that it is conceivable that model predictions will improve when the present linear networks are replaced by deep learning networks when presented with the same input and output representations used here (see Zhao et al. [171], for a discussion of loss functions for deep learning targeting numeric instead of

categorical output vectors). What is unclear is whether such networks will improve our understanding—the current linear networks offer the analyst connection weights between input and output representations that are straightforwardly linguistically interpretable. Where we think deep learning networks really will come into their own is bridging the gap between for instance the cochlea and retina on the one hand, and the heuristic representations (letter trigrams) that we have used as a starting point for our functional theory of the mappings between form and meaning.

Network flexibility. An important challenge for deep learning networks designed to model human lexical processing is to keep the networks flexible and open to rapid updating. This openness to continuous learning is required not only by findings on categorization [40, 41, 41, 42], including phonetic categorization [172, 173], but is also visible in time series of reaction times. Baayen and Hendrix [17] reported improved prediction of visual lexical decision reaction times in the British Lexicon Project when a naive discriminative learning network is updated from trial to trial, simulating the within-experiment learning that goes on as subjects proceed through the experiment. In natural speech, the phenomenon of phonetic convergence between interlocutors [174] likewise bears witness to a lexicon that is constantly recalibrated in interaction with its environment (for the many ways in which interlocutors' speech aligns, see [175]). Likewise, the rapidity with which listeners adapt to foreign accents [176], within a few minutes, shows that lexical networks exhibit fast local optimization. As deep learning networks typically require huge numbers of cycles through the training data, once trained, they are at risk of not having the required flexibility for fast local adaptation. This risk is reduced for (incremental) wide learning, as illustrated by the phonetic convergence evidenced by the model of Arnold et al. [18] for auditory comprehension (a very similar point was made by Levelt [177] in his critique of connectionist models in the nineties.).

Open questions. This initial study necessarily leaves many questions unanswered. Compounds and inflected derived words have not been addressed, and morphological operations such as reduplication, infixation, and nonconcatenative morphology provide further testing grounds for the present approach. We also need to address the processing of compounds with multiple constituents, ubiquitous in languages as different as German, Mandarin, and Estonian. Furthermore, in actual utterances, words undergo assimilation at their boundaries, and hence a crucial test case for the discriminative lexicon is to model the production and comprehension of words in multi-word utterances. For our production model, we have also completely ignored stress, syllabification and pitch, which however likely has rendered the production task more difficult than necessary.

A challenge for discriminative morphology is the development of proper semantic matrices. For languages with rich morphology, such as Estonian, current morphological parsers will identify case-inflected words as nominatives, genitives, or partitives (among others), but these labels for inflectional forms do not correspond to the semantic functions encoded in these forms (see, e.g., [178, 179]). What is required are computational tools that detect the appropriate inflectional lexomes for these words in the sentences or utterances in which they are embedded.

A related problem specifically for the speech production model is how to predict the strongly reduced word forms that are rampant in conversational speech [111]. Our auditory comprehension network $\mathbf{F}_a$ is confronted with reduced forms in training, and the study of Arnold et al. [18] indicates that present identification accuracy may approximate human performance for single-word recognition. Whereas for auditory comprehension we are making progress, what is required for speech production is a much more detailed understanding of the circumstances under which speakers actually use reduced word forms. The factors driving reduction may be in part captured by the semantic vectors, but we anticipate that aspects of the unfolding discourse play an important role as well, which brings us back to the unresolved question of how to best model not isolated words but words in context.

An important challenge for our theory, which formalizes incremental implicit learning, is how to address and model the wealth of explicit knowledge that speakers have about their language. We play with the way words sound in poetry; we are perplexed by words' meanings and reinterpret them so that they make more sense (e.g., the folk etymology [180] reflected in modern English *crawfish*, a crustacean and not a fish, originally Middle English *crevis*; compare *écrevisse* in French), we teach morphology in schools, and we even find that making morphological relations explicit may be beneficial for aphasic patients [181]. The rich culture of word use cannot but interact with the implicit system, but how exactly this interaction unfolds and what its consequences are at both sides of the conscious/unconscious divide is presently unclear.

Although many questions remain, our model is remarkably successful in capturing many aspects of lexical and morphological processing in both comprehension and production. Since the model builds on a combination of linguistically motivated 'smart' low-level representations for form and meaning and large but otherwise straightforward and mathematically well-understood two-layer linear networks (to be clear, networks without any hidden layers), the conclusion seems justified that, algorithmically, the "mental lexicon" may be much simpler than previously thought.

Appendix

A. Vectors, Matrices, and Matrix Multiplication

Figure 12 presents two data points, *a* and *b*, with coordinates (3, 4) and (-2, -3). Arrows drawn from the origin to these points are shown in blue with a solid and dashed line, respectively. We refer to these points as vectors. Vectors are denoted with lower case letters in bold, and we place the x coordinate of a point next to the y coordinate:

$$\begin{aligned} \mathbf{a} &= \begin{pmatrix} 3 & 4 \end{pmatrix}, \\ \mathbf{b} &= \begin{pmatrix} -2 & -3 \end{pmatrix}. \end{aligned} \tag{A.1}$$

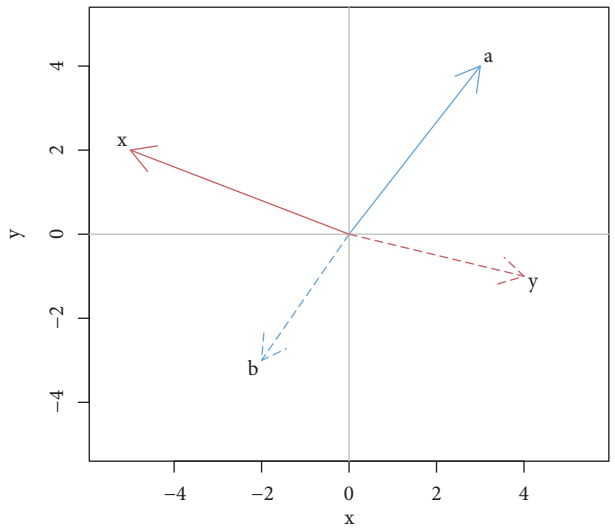


FIGURE 12: Points a and b can be transformed into points p and q by a linear transformation.

We can represent **a** and **b** jointly by placing them next to each other in a matrix. For matrices, we use capital letters in bold font.

$$\mathbf{A} = \begin{pmatrix} 3 & 4 \\ -2 & -3 \end{pmatrix}. \tag{A.2}$$

In addition to the points a and b, Figure 12 also shows two other datapoints, x and y. The matrix for these two points is

$$\mathbf{B} = \begin{pmatrix} -5 & 2 \\ 4 & -1 \end{pmatrix}. \tag{A.3}$$

Let us assume that points a and b represent the forms of two words ω_1 and ω_2 and that the points x and y represent the meanings of these two words (below, we discuss in detail how exactly words' forms and meanings can be represented as vectors of numbers). As morphology is the study of the relation between words' forms and their meanings; we are interested in how to transform points a and b into points x and y and, likewise, in how to transform points x and y back into points a and b. The first transformation is required for comprehension and the second transformation for production.

We begin with the transformation for comprehension. Formally, we have matrix **A**, which we want to transform into matrix **B**. A way of mapping **A** onto **B** is by means of a linear transformation using matrix multiplication. For 2×2 matrices, matrix multiplication is defined as follows:

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \begin{pmatrix} a_1 x_1 + a_2 y_1 & a_1 x_2 + a_2 y_2 \\ b_1 x_1 + b_2 y_1 & b_1 x_2 + b_2 y_2 \end{pmatrix}. \tag{A.4}$$

Such a mapping of **A** onto **B** is given by a matrix **F**,

$$\mathbf{F} = \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix}, \tag{A.5}$$

and it is straightforward to verify that indeed

$$\begin{pmatrix} 3 & 4 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} -5 & 2 \\ 4 & -1 \end{pmatrix}. \tag{A.6}$$

Using matrix notation, we can write

$$\mathbf{AF} = \mathbf{B}. \tag{A.7}$$

Given **A** and **B**, how do we obtain **F**? The answer is straightforward, but we need two further concepts: the identity matrix and the matrix inverse. For multiplication of numbers, the identity multiplication is multiplication with 1. For matrices, multiplication with the identity matrix **I** leaves the locations of the points unchanged. The identity matrix has ones on the main diagonal and zeros elsewhere. It is easily verified that indeed

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}. \tag{A.8}$$

For numbers, the inverse of multiplication with s is dividing by s:

$$\left(\frac{1}{s}\right)(sx) = x. \tag{A.9}$$

For matrices, the inverse of a square matrix **X** is that matrix $\mathbf{X}^{-1}$ such that their product is the identity matrix:

$$\mathbf{X}^{-1}\mathbf{X} = \mathbf{X}\mathbf{X}^{-1} = \mathbf{I}. \tag{A.10}$$

For nonsquare matrices, the inverse $\mathbf{Y}^{-1}$ is defined such that

$$\mathbf{Y}^{-1}(\mathbf{YX}) = (\mathbf{XY})\,Y^{-1} = \mathbf{X}. \tag{A.11}$$

We find the square matrix **F** that maps the square matrices **A** onto **B** as follows:

$$\begin{aligned} \mathbf{AF} &= \mathbf{B} \\ \mathbf{A}^{-1}\mathbf{AF} &= \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{IF} &= \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{F} &= \mathbf{A}^{-1}\mathbf{B}. \end{aligned} \tag{A.12}$$

Since for the present example **A** and its inverse happen to be identical ($\mathbf{A}^{-1} = \mathbf{A}$), we obtain (A.6).

F maps words' form vectors onto words' semantic vectors. Let us now consider the reverse and see how we can obtain words' form vectors from words' semantic vectors. That is, given **B**, we want to find that matrix **G** which maps **B** onto **A**, i.e.,

$$\mathbf{BG} = \mathbf{A}. \tag{A.13}$$

Solving this equation proceeds exactly as above

$$\begin{aligned} \mathbf{BG} &= \mathbf{A} \\ \mathbf{B}^{-1}\mathbf{BG} &= \mathbf{B}^{-1}\mathbf{A} \\ \mathbf{IG} &= \mathbf{B}^{-1}\mathbf{A} \\ \mathbf{G} &= \mathbf{B}^{-1}\mathbf{A}. \end{aligned} \tag{A.14}$$

The inverse of **B** is

$$\mathbf{B}^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{4}{3} & \frac{5}{3} \end{pmatrix}, \tag{A.15}$$

and hence we have

$$\begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{4}{3} & \frac{5}{3} \end{pmatrix} \begin{pmatrix} 3 & 4 \\ -2 & -3 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} = \mathbf{G}. \tag{A.16}$$

The inverse of a matrix needs not exist. A square matrix that does not have an inverse is referred to as a singular matrix. Almost all matrices with which we work in the remainder of this study are singular. For singular matrices, an approximation of the inverse can be used, such as the Moore-Penrose generalized inverse (In R, the generalized inverse is available in the **MASS** package, function `ginv`. The numeric library that is used by `ginv` is LAPACK, available at `http://www.netlib.org/lapack`.). In this study, we denote the generalized inverse of a matrix **X** by $\mathbf{X}'$.

The examples presented here are restricted to 2×2 matrices, but the mathematics generalize to matrices of larger dimensions. When multiplying two matrices **X** and **Y**, the only constraint is that the number of columns of **X** is the same as the number of rows of **Y**. The resulting matrix has the same number of rows as **X** and the same number of columns as **Y**.

In the present study, the rows of matrices represent words and columns the features of these words. For instance, the row vectors of **A** can represent words' form features and the row vectors of and **B** the corresponding semantic features. The first row vector of **A** is mapped onto the first row vector of **B** as follows:

$$\begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} -5 & 2 \end{pmatrix} \tag{A.17}$$

A transformation matrix such as **F** can be represented as a two-layer network. The network corresponding to **F** is shown in Figure 13. When the input vector $(3, 4)$ is presented to the network, the nodes i_1 and i_2 are set to 3 and 4, respectively. To obtain the activations $o_1 = -5$ and $o_2 = 2$ for the output vector $(-5, 2)$, we cumulate the evidence at the input, weighted by the connection strengths specified on the edges of the graph. Thus, for o_1, we obtain the value $3 \times 1 + 4 \times -2 = -5$ and, for o_2, we have $3 \times 2 + 4 \times -1 = 2$. In other words, the network maps the input vector $(3, 4)$ onto the output vector $(-5, 2)$.

An important property of linear maps is that they are productive. We can present a novel form vector, say,

$$\mathbf{s} = \begin{pmatrix} 2 & 2 \end{pmatrix}, \tag{A.18}$$

to the network, and it will map this vector onto a novel semantic vector. Using matrices, this new semantic vector is straightforwardly calculated:

$$\begin{pmatrix} 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} -2 & 2 \end{pmatrix}. \tag{A.19}$$

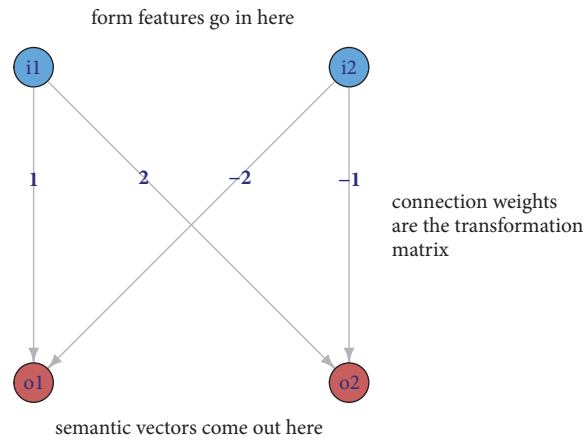


Figure 13: The linear network corresponding to the transformation matrix **F**. There are two input nodes, i_1 and i_2 (in blue),and two output nodes, o_1 and o_2, in red. When the input vector $(3, 4)$ is presented to the network, the value at o_1 is $3 \times 1 + 4 \times -2 = -5$ and that at o_2 is $3 \times 2 + 4 \times -1 = 2$. The network maps the point $(3, 4)$ onto the point $(-5, 2)$.

B. Graph-Based Triphone Sequencing

The hypothesis underlying graph-based triphone sequencing is that, in the directed graph that has triphones as vertices and an edge between any two vertices that overlap (i.e., segments 2 and 3 of triphone 1 are identical to segments 1 and 2 of triphone 2), the path from a word's initial triphone (the triphone starting with #) to a word's final triphone (the triphone ending with a #) is the path receiving the best support of all possible paths from the initial to the final triphone. Unfortunately, the directed graph containing all vertices and edges is too large to make computations tractable in a reasonable amount of computation time. The following algorithm is a heuristic algorithm that first collects potentially relevant vertices and edges and then calculates all paths starting from any initial vertex, using the `all_simple_paths` function from the **igraph** package [127]. As a first step, all vertices that are supported by the stem and whole word with network support exceeding 0.1 are selected. To the resulting set, those vertices are added that are supported by the affix, which is conditional on these vertices having a network support exceeding 0.95. A directed graph can now be constructed and, for each initial vertex, all paths starting at these vertices are calculated. The subset of those paths reaching a final vertex is selected, and the support for each path is calculated. As longer paths trivially will have greater support, the support for a path is weighted for its length, simply by dividing the raw support by the path length. Paths are ranked by weighted path support, and the path with maximal support is selected as the acoustic image driving articulation.

It is possible that no path from an initial vertex to a final vertex is found, due to critical boundary triphones not being instantiated in the data on which the model was trained. This happens when the model is assessed on novel inflected

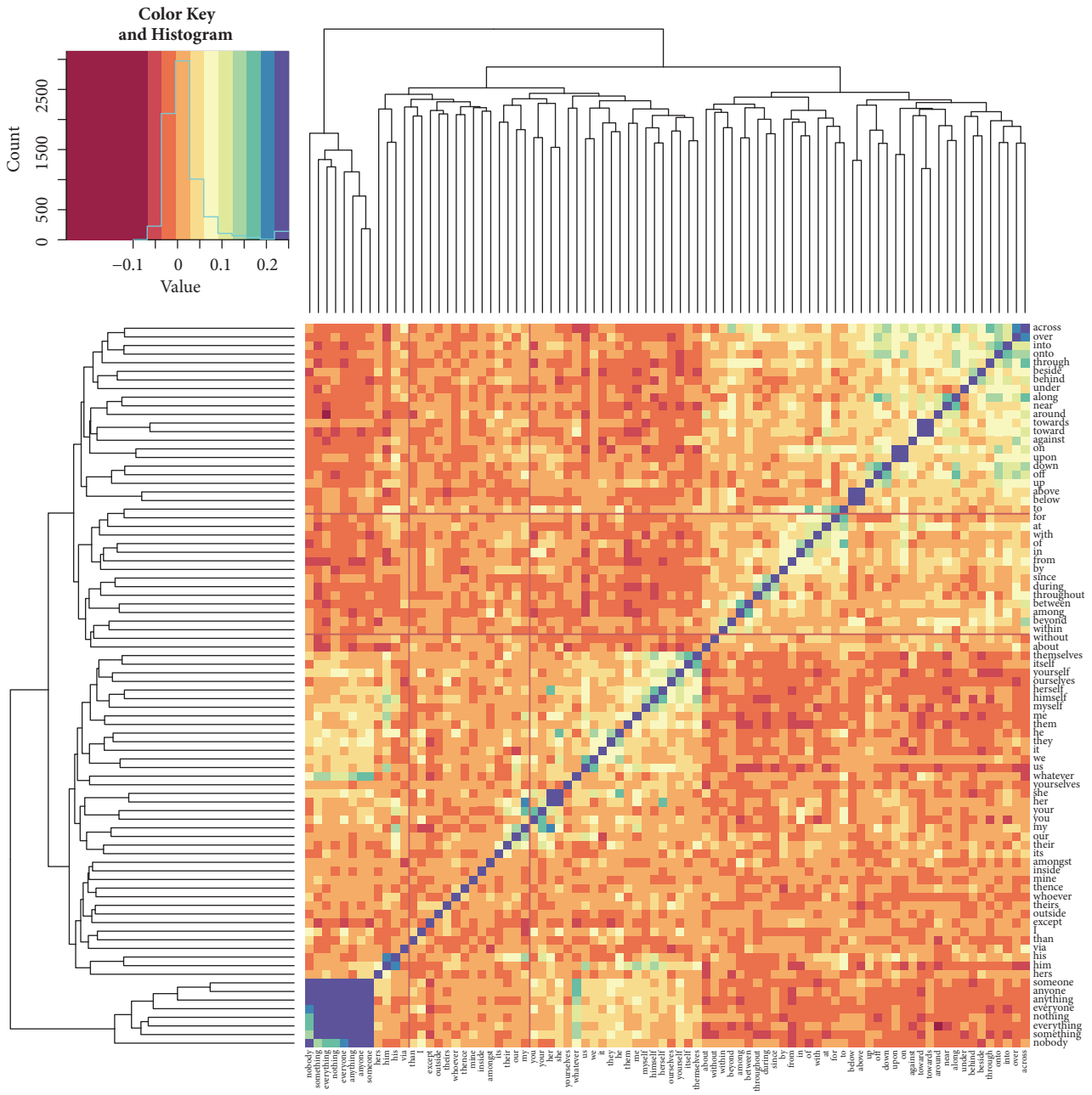


Figure 14: Heatmap for the correlations of the semantic vectors of pronouns and prepositions. Note that quantifier pronouns form a group of their own that prepositions and the other pronouns form distinguisable groups and that, within groups, further subclusters are visible, e.g., for *we* and *us*, *she* and *her*, and *across* and *over*. All diagonal elements represent correlations equal to 1 (this is not brought out by the heatmap, which color-codes diagonal elements with the color for the highest correlation of the range we specified, 0.25).

and derived words under cross-validation. Therefore, vertices and edges are added to the graph for all nonfinal vertices that are at the end of any of the paths starting at any initial triphone. Such novel vertices and edges are assigned zero network support. Typically, paths from the initial vertex to the final vertex can now be constructed, and path support is evaluated as above.

For an algorithm that allows triphone paths to include cycles, which may be the case in languages with much richer morphology than English, see Baayen et al. [85], and for code the R package **WpmWithLdl** described therein.

C. A Heatmap for Function Words (Pronouns and Prepositions)

The heatmap for the correlations of the semantic vectors of pronouns and prepositions is presented in Figure 14.

Data Availability

With the exception of the data from the Distributed Little Red Hen Lab, all primary data are publicly available from the cited sources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are indebted to Geoff Hollis, Jessie Nixon, Chris Westbury, and Luis Mienhardt for their constructive feedback on earlier versions of this manuscript. This research was supported by an ERC advanced Grant (no. 742545) to the first author.

References

[1] G. Frege, "Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought," in *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, pp. 1–82, 1967.

[2] B. Russell, *An Inquiry into Meaning and Truth*, Allen and Unwin, London, UK, 1942.

[3] R. Montague, "The proper treatment of quantification in ordinary english," in *Approaches to Natural Language*, pp. 221–242, Springer, 1973.

[4] N. Chomsky, *Syntactic Structures*, vol. 33, Mouton, The Hague, 1957.

[5] W. J. M. Levelt, *Formal Grammars in Linguistics And Psycholinguistics: Volume 1: An Introduction to the Theory of Formal Languages and Automata, Volume 2: Applications in Linguistic Theory; Volume 3: Psycholinguistic Applications*, John Benjamins Publishing, 2008.

[6] D. F. Kleinschmidt and T. Florian Jaeger, "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," *Psychological Review*, vol. 122, no. 2, pp. 148–203, 2015.

[7] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[8] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: illustrations, sources and a solution," *BMC Bioinformatics*, vol. 8, 2007.

[9] A. Hannun, C. Case, J. Casper et al., "Deep speech: Scaling up end-to-end speech recognition," 2014, https://arxiv.org/abs/1412.5567.

[10] Z. Tüske, P. Golik, R. Schlüter, and H. Ney, "Acoustic modeling with deep neural networks using raw time signal for lvcsr," in *Proceedings of the INTERSPEECH*, pp. 890–894, 2014.

[11] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[12] T. Hannagan, J. C. Ziegler, S. Dufau, J. Fagot, and J. Grainger, "Deep learning of orthographic representations in baboons," *PLoS ONE*, vol. 9, no. 1, 2014.

[13] J. Grainger, S. Dufau, M. Montant, J. C. Ziegler, and J. Fagot, "Orthographic processing in baboons (Papio papio)," *Science*, vol. 336, no. 6078, pp. 245–248, 2012.

[14] D. Scarf, K. Boy, A. U. Reinert, J. Devine, O. Güntürkün, and M. Colombo, "Orthographic processing in pigeons (columba livia)," *Proceedings of the National Academy of Sciences*, vol. 113, no. 40, pp. 11272–11276, 2016.

[15] M. Linke, F. Bröker, M. Ramscar, and R. H. Baayen, "Are baboons learning "orthographic" representations? Probably not," *PLoS ONE*, vol. 12, no. 8, 2017.

[16] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, pp. 106–154, 1962.

[17] R. H. Baayen and P. Hendrix, "Two-layer networks, non-linear separation, and human learning," in *From Semantics to Dialectometry. Festschrift in honor of John Nerbonne. Tributes 32*, M. Wieling, M. Kroon, G. Van Noord, and G. Bouma, Eds., pp. 13–22, College Publications, 2017.

[18] D. Arnold, F. Tomaschek, K. Sering, F. Lopez, and R. H. Baayen, "Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit," *PLoS ONE*, vol. 12, no. 4, 2017.

[19] P. Zwitserlood, "Processing and representation of morphological complexity in native language comprehension and production," in *The Construction of Words*, G. E. Booij, Ed., pp. 583–602, Springer, 2018.

[20] G. Booij, "Construction morphology," *Language and Linguistics Compass*, vol. 4, no. 7, pp. 543–555, 2010.

[21] J. P. Blevins, "Word-based morphology," *Journal of Linguistics*, vol. 42, no. 3, pp. 531–573, 2006.

[22] G. Stump, *Inflectional Morphology: A Theory of Paradigm Structure*, Cambridge University Press, 2001.

[23] R. Beard, "On the extent and nature of irregularity in the lexicon," *Lingua*, vol. 42, no. 4, pp. 305–341, 1977.

[24] P. H. Matthews, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, Cambridge, 1974.

[25] P. H. Matthews, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, Cambridge, 1991.

[26] C. Hockett, "The origin of speech," *Scientific American*, vol. 203, pp. 89–97, 1960.

[27] M. Erelt, *Estonian Language*, M. Erelt, Ed., Estonian Academy Publishers, Tallinn, Estonia, 2003.

[28] P. H. Matthews, *Grammatical Theory in the United States from Bloomfield to Chomsky*, Cambridge University Press, Cambridge, 1993.

[29] N. Chomsky and M. Halle, *The Sound Pattern of English*, Harper and Row, NY, USA, 1968.

[30] T. Landauer and S. Dumais, "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211–240, 1997.

[31] W. Weaver, "Translation," in *Machine Translation of Languages: Fourteen Essays*, W. N. Locke and A. D. Booth, Eds., pp. 15–23, MIT Press, Cambridge, 1955.

[32] J. R. Firth, *Selected Papers of J R Firth, 1952-59*, Indiana University Press, 1968.

[33] K. Lund and C. Burgess, "Producing high-dimensional semantic spaces from lexical co-occurrence," *Behavior Research Methods, Instruments, and Computers*, vol. 28, no. 2, pp. 203–208, 1996.

[34] C. Shaoul and C. Westbury, "Exploring lexical co-occurrence space using HiDEx," *Behavior Research Methods*, vol. 42, no. 2, pp. 393–413, 2010.

[35] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, pp. 3111–3119, 2013.

[36] A. Roelofs, "The WEAVER model of word-form encoding in speech production," *Cognition*, vol. 64, no. 3, pp. 249–284, 1997.

[37] W. J. M. Levelt, A. Roelofs, and A. S. Meyer, "A theory of lexical access in speech production," *Behavioral and Brain Sciences*, vol. 22, no. 1, pp. 1–38, 1999.

[38] M. Taft, "Interactive-activation as a Framework for Understanding Morphological Processing," *Language and Cognitive Processes*, vol. 9, no. 3, pp. 271–294, 1994.

[39] R. Schreuder and R. H. Baayen, "Modeling morphological processing," in *Morphological Aspects of Language Processing*, L. B. Feldman, Ed., pp. 131–154, Lawrence Erlbaum, Hillsdale, New Jersey, USA, 1995.

[40] B. C. Love, D. L. Medin, and T. M. Gureckis, "SUSTAIN: A Network Model of Category Learning," *Psychological Review*, vol. 111, no. 2, pp. 309–332, 2004.

[41] C. J. Marsolek, "What antipriming reveals about priming," *Trends in Cognitive Sciences*, vol. 12, no. 5, pp. 176–181, 2008.

[42] M. Ramscar and R. Port, "Categorization (without categories)," in *Handbook of Cognitive Linguistics*, E. Dabrowska and D. Divjak, Eds., pp. 75–99, De Gruyter, Berlin, Germany, 2015.

[43] R. A. Rescorla and A. R. Wagner, "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement," in *Classical Conditioning II: Current Research and Theory*, A. H. Black and W. F. Prokasy, Eds., pp. 64–99, Appleton Century Crofts, NY, USA, 1972.

[44] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *1960 WESCON Convention Record Part IV*, pp. 96–104, 1960.

[45] R. H. Baayen, R. Schreuder, and R. Sproat, "Morphology in the Mental Lexicon: A Computational Model for Visual Word Recognition," in *Lexicon Development for Speech and Language Processing*, F. van Eynde and D. Gibbon, Eds., pp. 267–291, Kluwer Academic Publishers, 2000.

[46] M. Taft, "A morphological-decomposition model of lexical representation," *Linguistics*, vol. 26, no. 4, pp. 657–667, 1988.

[47] M. W. Harm and M. S. Seidenberg, "Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes," *Psychological Review*, vol. 111, no. 3, pp. 662–720, 2004.

[48] P. C. Trimmer, J. M. McNamara, A. Houston, and J. A. R. Marshall, "Does natural selection favour the Rescorla-Wagner rule?" *Journal of Theoretical Biology*, vol. 302, pp. 39–52, 2012.

[49] R. R. Miller, R. C. Barnet, and N. J. Grahame, "Assessment of the Rescorla-Wagner model," *Psychological Bulletin*, vol. 117, no. 3, pp. 363–386, 1995.

[50] L. J. Kamin, "Predictability, surprise, attention, and conditioning," in *Punishment and Aversive Behavior*, B. A. Campbell and R. M. Church, Eds., pp. 276–296, Appleton-Century-Crofts, NY, USA, 1969.

[51] R. A. Rescorla, "Pavlovian conditioning. It's not what you think it is," *American Psychologist*, vol. 43, no. 3, pp. 151–160, 1988.

[52] M. Ramscar, D. Yarlett, M. Dye, K. Denny, and K. Thorpe, "The Effects of Feature-Label-Order and Their Implications for Symbolic Learning," *Cognitive Science*, vol. 34, no. 6, pp. 909–957, 2010.

[53] M. Ramscar and D. Yarlett, "Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition," *Cognitive Science*, vol. 31, no. 6, pp. 927–960, 2007.

[54] R. H. Baayen, P. Milin, D. F. Đurđević, P. Hendrix, and M. Marelli, "An Amorphous Model for Morphological Processing in Visual Comprehension Based on Naive Discriminative Learning," *Psychological Review*, vol. 118, no. 3, pp. 438–481, 2011.

[55] F. M. Del Prado Martín, R. Bertram, T. Häikiö, R. Schreuder, and R. H. Baayen, "Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew," *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 30, pp. 1271–1278, 2004.

[56] P. Milin, D. F. Durdević, and F. M. Del Prado Martín, "The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian," *Journal of Memory and Language*, vol. 60, no. 1, pp. 50–64, 2009.

[57] B. K. Bergen, "The psychological reality of phonaesthemes," *Language*, vol. 80, no. 2, pp. 290–311, 2004.

[58] P. Milin, L. B. Feldman, M. Ramscar, P. Hendrix, and R. H. Baayen, "Discrimination in lexical decision," *PLoS ONE*, vol. 12, no. 2, 2017.

[59] R. H. Baayen, P. Milin, and M. Ramscar, "Frequency in lexical processing," *Aphasiology*, vol. 30, no. 11, pp. 1174–1220, 2016.

[60] M. Marelli and M. Baroni, "Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics," *Psychological Review*, vol. 122, no. 3, pp. 485–515, 2015.

[61] T. Sering, P. Milin, and R. H. Baayen, "Language comprehension as a multiple label classification problem," *Statistica Neerlandica*, pp. 1–15, 2018.

[62] R. Kaye and R. Wilson, *Linear Algebra*, Oxford University Press, 1998.

[63] S. H. Ivens and B. L. Koslin, *Demands for Reading Literacy Require New Accountability Methods*, Touchstone Applied Science Associates, 1991.

[64] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Processes*, vol. 25, no. 2-3, pp. 259–284, 1998.

[65] H. Schmid, "Improvements in Part-of-Speech Tagging with an Application to German," in *Proceedings of the ACL SIGDAT-Workshop*, pp. 13–25, Dublin, Ireland, 1995.

[66] R. H. Baayen, R. Piepenbrock, and L. Gulikers, *The CELEX Lexical Database (CD-ROM)*, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA, 1995.

[67] J. Mitchell and M. Lapata, "Vector-based models of semantic composition," in *Proceedings of the ACL*, pp. 236–244, 2008.

[68] A. Lazaridou, M. Marelli, R. Zamparelli, and M. Baroni, "Compositionally derived representations of morphologically complex words in distributional semantics," in *Proceedings of the ACL (1)*, pp. 1517–1526, 2013.

[69] R. Cotterell, H. Schütze, and J. Eisner, "Morphological smoothing and extrapolation of word embeddings," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 1651–1660, 2016.

[70] B. D. Zeller, S. Pado, and J. Šnajder, "Towards semantic validation of a derivational lexicon," in *Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 1728–1739, 2014.

[71] H. Marchand, *The Categories and Types of Present-Day English Word Formation. A Synchronic-Diachronic Approach*, Beck'sche Verlagsbuchhandlung, München, Germany, 1969.

[72] L. Bauer, *English Word-Formation*, Cambridge University Press, Cambridge, 1983.

[73] I. Plag, *Word-Formation in English*, Cambridge University Press, Cambridge, 2003.

[74] R. Beard, *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*, State University of New York Press, Albany, NY, USA, 1995.

[75] K. Geeraert, J. Newman, and R. H. Baayen, "Idiom Variation: Experimental Data and a Blueprint of a Computational Model," *Topics in Cognitive Science*, vol. 9, no. 3, pp. 653–669, 2017.

[76] C. Shaoul, S. Bitschau, N. Schilling et al., "ndl2: Naive discriminative learning: an implementation in R. R package," 2015.

[77] P. Milin, D. Divjak, and R. H. Baayen, "A learning perspective on individual differences in skilled reading: Exploring and exploiting orthographic and semantic discrimination cues," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 43, no. 11, pp. 1730–1751, 2017.

[78] M. Ramscar, P. Hendrix, C. Shaoul, P. Milin, and R. H. Baayen, "Nonlinear dynamics of lifelong learning: the myth of cognitive decline," *Topics in Cognitive Science*, vol. 6, pp. 5–42, 2014.

[79] M. Ramscar, C. C. Sun, P. Hendrix, and R. H. Baayen, "The Mismeasurement of Mind: Life-Span Changes in Paired-Associate-Learning Scores Reflect the "Cost" of Learning, Not Cognitive Decline," *Psychological Science*, vol. 28, no. 8, pp. 1171–1179, 2017.

[80] G. desRosiers and D. Ivison, "Paired associate learning: Form 1 and form 2 of the Wechsler memory scale," *Archives of Clinical Neuropsychology*, vol. 3, no. 1, pp. 47–67, 1988.

[81] E. Bruni, N. K. Tran, and M. Baroni, "Multimodal distributional semantics," *Journal of Artificial Intelligence Research*, vol. 49, pp. 1–47, 2014.

[82] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-PLUS*, Springer, NY, USA, 2002.

[83] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behavior Research Methods*, vol. 45, no. 4, pp. 1191–1207, 2013.

[84] D. Kastovsky, "Productivity in word formation," *Linguistics*, vol. 24, no. 3, pp. 585–600, 1986.

[85] R. H. Baayen, Y. Chuang, and J. P. Blevins, "Inflectional morphology with linear mappings," *The Mental Lexicon*, vol. 13, no. 2, pp. 232–270, 2018.

[86] Burnage, *CELEX; A Guide for Users*, Centre for Lexical Information, Nijmegen, The Netherlands, 1988.

[87] C. McBride-Chang, "Models of Speech Perception and Phonological Processing in Reading," *Child Development*, vol. 67, no. 4, pp. 1836–1856, 1996.

[88] G. C. Van Orden, H. Kloos et al., "The question of phonology and reading," in *The Science of Reading: A Handbook*, pp. 61–78, 2005.

[89] D. Jared and K. O'Donnell, "Skilled adult readers activate the meanings of high-frequency words using phonology: Evidence from eye tracking," *Memory & Cognition*, vol. 45, no. 2, pp. 334–346, 2017.

[90] G. S. Dell, "A Spreading-Activation Theory of Retrieval in Sentence Production," *Psychological Review*, vol. 93, no. 3, pp. 283–321, 1986.

[91] A. Marantz, "No escape from morphemes in morphological processing," *Language and Cognitive Processes*, vol. 28, no. 7, pp. 905–916, 2013.

[92] M. Bozic, W. D. Marslen-Wilson, E. A. Stamatakis, M. H. Davis, and L. K. Tyler, "Differentiating morphology, form, and meaning: Neural correlates of morphological complexity," *Cognitive Neuroscience*, vol. 19, no. 9, pp. 1464–1475, 2007.

[93] R. F. Port and A. P. Leary, "Against formal phonology," *Language*, vol. 81, no. 4, pp. 927–964, 2005.

[94] J. P. Blevins, "Stems and paradigms," *Language*, vol. 79, no. 4, pp. 737–767, 2003.

[95] R. Forsyth and D. Holmes, "Feature-finding for test classification," *Literary and Linguistic Computing*, vol. 11, no. 4, pp. 163–174, 1996.

[96] H. Pham and R. H. Baayen, "Vietnamese compounds show an anti-frequency effect in visual lexical decision," *Language, Cognition, and Neuroscience*, vol. 30, no. 9, pp. 1077–1095, 2015.

[97] L. Cohen and S. Dehaene, "Ventral and dorsal contributions to word reading," in *The Cognitive Neurosciences*, M. S. Gazzaniga, Ed., pp. 789–804, Massachusetts Institute of Technology, 2009.

[98] E. Keuleers, P. Lacey, K. Rastle, and M. Brysbaert, "The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words," *Behavior Research Methods*, vol. 44, no. 1, pp. 287–304, 2012.

[99] R. K. Wagner, J. K. Torgesen, and C. A. Rashotte, "Development of reading-related phonological processing abilities: new evidence of bidirectional causality from a latent variable longitudinal study," *Developmental Psychology*, vol. 30, no. 1, pp. 73–87, 1994.

[100] K. F. E. Wong and H.-C. Chen, "Orthographic and phonological processing in reading Chinese text: Evidence from eye fixations," *Language and Cognitive Processes*, vol. 14, no. 5-6, pp. 461–480, 1999.

[101] R. L. Newman, D. Jared, and C. A. Haigh, "Does phonology play a role when skilled readers read high-frequency words? Evidence from ERPs," *Language and Cognitive Processes*, vol. 27, no. 9, pp. 1361–1384, 2012.

[102] D. Jared, J. Ashby, S. J. Agauas, and B. A. Levy, "Phonological activation of word meanings in grade 5 readers," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 42, no. 4, pp. 524–541, 2016.

[103] T. Bitan, A. Kaftory, A. Meiri-Leib, Z. Eviatar, and O. Peleg, "Phonological ambiguity modulates resolution of semantic ambiguity during reading: An fMRI study of Hebrew," *Neuropsychology*, vol. 31, no. 7, pp. 759–777, 2017.

[104] D. Jared and S. Bainbridge, "Reading homophone puns: Evidence from eye tracking," *Canadian Journal of Experimental Psychology*, vol. 71, no. 1, pp. 2–13, 2017.

[105] S. Amenta, M. Marelli, and S. Sulpizio, "From sound to meaning: Phonology-to-Semantics mapping in visual word recognition," *Psychonomic Bulletin & Review*, vol. 24, no. 3, pp. 887–893, 2017.

[106] M. Perrone-Bertolotti, J. Kujala, J. R. Vidal et al., "How silent is silent reading? intracerebral evidence for top-down activation of temporal voice areas during reading," *The Journal of Neuroscience*, vol. 32, no. 49, pp. 17554–17562, 2012.

[107] B. Yao, P. Belin, and C. Scheepers, "Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex," *Journal of Cognitive Neuroscience*, vol. 23, no. 10, pp. 3146–3152, 2011.

[108] M. Coltheart, B. Curtis, P. Atkins, and M. Haller, "Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches," *Psychological Review*, vol. 100, no. 4, pp. 589–608, 1993.

[109] M. Coltheart, "Modeling reading: The dual-route approach," in *The Science of Reading: A Handbook*, pp. 6–23, 2005.

[110] M. Pitt, K. Johnson, E. Hume, S. Kiesling, and W. Raymond, "The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability," *Speech Communication*, vol. 45, no. 1, pp. 89–95, 2005.

[111] K. Johnson, "Massive reduction in conversational American English," in *Proceedings of the Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, pp. 29–54, The National International Institute for Japanese Language, Tokyo, Japan, 2004.

[112] C. Phillips, "Levels of representation in the electrophysiology of speech perception," *Cognitive Science*, vol. 25, no. 5, pp. 711–731, 2001.

[113] R. L. Diehl, A. J. Lotto, and L. L. Holt, "Speech perception," *Annual Review of Psychology*, vol. 55, pp. 149–179, 2004.

[114] D. Norris and J. M. McQueen, "Shortlist B: a Bayesian model of continuous speech recognition," *Psychological Review*, vol. 115, no. 2, pp. 357–395, 2008.

[115] S. Hawkins, "Roles and representations of systematic fine phonetic detail in speech understanding," *Journal of Phonetics*, vol. 31, no. 3-4, pp. 373–405, 2003.

[116] C. Cucchiarini and H. Strik, "Automatic Phonetic Transcription: An overview," in *Proceedings of the 15th ICPhS*, pp. 347–350, Barcelona, Spain, 2003.

[117] K. Johnson, *The Auditory/Perceptual Basis for Speech Segmentation*, Ohio State University Working Papers in Linguistics, 1997.

[118] H. Fletcher, "Auditory patterns," *Reviews of Modern Physics*, vol. 12, no. 1, pp. 47–65, 1940.

[119] D. Arnold, "Acousticndlcoder: Coding sound files for use with ndl. R package version 1.0.1," 2017.

[120] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016.

[121] J. Pickett and I. Pollack, "Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt," *Language and Speech*, vol. 6, pp. 151–164, 1963.

[122] L. Shockey, "Perception of reduced forms by non-native speakers of English," in *Sound Patterns of Spontaneous Speech*, D. Duez, Ed., pp. 97–100, ESCA, Aix, 1998.

[123] M. Ernestus, H. Baayen, and R. Schreuder, "The recognition of reduced word forms," *Brain and Language*, vol. 81, no. 1-3, pp. 162–173, 2002.

[124] R. H. Baayen, C. Shaoul, J. Willits, and M. Ramscar, "Comprehension without segmentation: a proof of concept with naive discriminative learning," *Language, Cognition and Neuroscience*, vol. 31, no. 1, pp. 106–128, 2016.

[125] C. P. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3-4, pp. 155–180, 1992.

[126] J. A. Tourville and F. H. Guenther, "The DIVA model: A neural theory of speech acquisition and production," *Language and Cognitive Processes*, vol. 26, no. 7, pp. 952–981, 2011.

[127] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal, Complex Systems*, vol. 1695, no. 5, pp. 1–9, 2006.

[128] M. Ernestus and R. H. Baayen, "Predicting the unpredictable: Interpreting neutralized segments in Dutch," *Language*, vol. 79, no. 1, pp. 5–38, 2003.

[129] R. Weingarten, G. Nottbusch, and U. Will, "Morphemes, syllables and graphemes in written word production," in *Multidisciplinary Approaches to Speech Production*, T. Pechmann and C. Habel, Eds., pp. 529–572, Mouton de Gruyter, Berlin, Germany, 2004.

[130] R. Bertram, F. E. Tønnessen, S. Strömqvist, J. Hyönä, and P. Niemi, "Cascaded processing in written compound word production," *Frontiers in Human Neuroscience*, vol. 9, pp. 1–10, 2015.

[131] T. Cho, "Effects of morpheme boundaries on intergestural timing: Evidence from Korean," *Phonetica*, vol. 58, no. 3, pp. 129–162, 2001.

[132] S. N. Wood, *Generalized Additive Models*, Chapman & Hall/CRC, NY, USA, 2017.

[133] M. Seidenberg, "Sublexical structures in visual word recognition: Access units or orthographic redundancy," in *Attention and Performance XII*, M. Coltheart, Ed., pp. 245–264, Lawrence Erlbaum Associates, Hove, 1987.

[134] J. M. McQueen, "Segmentation of continuous speech using phonotactics," *Journal of Memory and Language*, vol. 39, no. 1, pp. 21–46, 1998.

[135] J. B. Hay, *Causes and Consequences of Word Structure*, Routledge, NY, USA, London, UK, 2003.

[136] J. B. Hay, "From speech perception to morphology: Affix-ordering revisited," *Language*, vol. 78, pp. 527–555, 2002.

[137] J. B. Hay and R. H. Baayen, "Phonotactics, parsing and productivity," *Italian Journal of Linguistics*, vol. 15, no. 1, pp. 99–130, 2003.

[138] J. Hay, J. Pierrehumbert, and M. Beckman, "Speech perception, well-formedness, and the statistics of the lexicon," *Papers in laboratory phonology VI*, pp. 58–74, 2004.

[139] M. Ferro, C. Marzi, and V. Pirrelli, "A self-organizing model of word storage and processing: implications for morphology learning," *Lingue e Linguaggio*, vol. 10, no. 2, pp. 209–226, 2011.

[140] F. Chersi, M. Ferro, G. Pezzulo, and V. Pirrelli, "Topological self-organization and prediction learning support both action and lexical chains in the brain," *Topics in Cognitive Science*, vol. 6, no. 3, pp. 476–491, 2014.

[141] V. Pirrelli, M. Ferro, and C. Marzi, "Computational complexity of abstractive morphology," in *Understanding and Measuring Morphological Complexity*, M. Bearman, D. Brown, and G. G. Corbett, Eds., pp. 141–166, Oxford University Press, 2015.

[142] C. Marzi, M. Ferro, and V. Pirrelli, "Is inflectional irregularity dysfunctional to human processing?" in *Abstract Booklet, The Mental Lexicon 2018*, V. Kuperman, Ed., University of Alberta, Edmonton, 2018.

[143] E. Keuleers, M. Stevens, P. Mandera, and M. Brysbaert, "Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment," *The Quarterly Journal of Experimental Psychology*, vol. 68, no. 8, pp. 1665–1692, 2015.

[144] M. Ernestus, *Voice Assimilation And Segment Reduction in Casual Dutch. A Corpus-Based Study of the Phonology-Phonetics Interface*, LOT, Utrecht, the Netherlands, 2000.

[145] R. Kemps, M. Ernestus, R. Schreuder, and H. Baayen, "Processing reduced word forms: The suffix restoration effect," *Brain and Language*, vol. 90, no. 1-3, pp. 117–127, 2004.

[146] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, June 2005.

[147] G. Hickok, "The architecture of speech production and the role of the phoneme in speech processing," *Language, Cognition and Neuroscience*, vol. 29, no. 1, pp. 2–20, 2014.

[148] B. V. Tucker, M. Sims, and R. H. Baayen, *Opposing Forces on Acoustic Duration*, Manuscript, University of Alberta and University of Tübingen, 2018.

[149] I. Plag, J. Homann, and G. Kunter, "Homophony and morphology: The acoustics of word-final S in English," *Journal of Linguistics*, vol. 53, no. 1, pp. 181–216, 2017.

[150] F. Tomaschek, I. Plag, M. Ernestus, and R. H. Baayen, *Modeling the Duration of Word-Final s in English with Naive Discriminative Learning*, Manuscript, University of Siegen/Tübingen/Nijmegen, 2018.

[151] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[152] B. Russell, "On denoting," *Mind*, vol. 14, no. 56, pp. 479–493, 1905.

[153] N. Hornstein, *Logical Form: From GB to Minimalism*, Blackwell, 1995.

[154] R. Jackendoff, *Semantic Structures*, MIT Press, Cambridge, 1990.

[155] J. J. McCarthy, "A prosodic theory of non-concatenative morphology," *Linguistic Inquiry*, vol. 12, pp. 373–418, 1981.

[156] A. Ussishkin, "A Fixed Prosodic Theory of Nonconcatenative Templaticmorphology," *Natural Language & Linguistic Theory*, vol. 23, no. 1, pp. 169–218, 2005.

[157] A. Ussishkin, "Affix-favored contrast inequity and psycholinguistic grounding for non-concatenative morphology," *Morphology*, vol. 16, no. 1, pp. 107–125, 2006.

[158] R. W. Young and W. Morgan, *The Navajo Language: A Grammar and Colloquial Dictionary*, University of New Mexico Press, 1980.

[159] S. C. Levinson and A. Majid, "The island of time: Yélî Dnye, the language of Rossel island," *Frontiers in psychology*, vol. 4, 2013.

[160] S. C. Levinson, "The language of space in Yélî Dnye," in *Grammars of Space: Explorations in Cognitive Diversity*, pp. 157–203, Cambridge University Press, 2006.

[161] E. Shafaei-Bajestan and R. H. Baayen, "Wide Learning for Auditory Comprehension," in *Proceedings of the Interspeech 2018*, 2018.

[162] J. Hay and K. Drager, "Stuffed toys and speech perception," *Linguistics*, vol. 48, no. 4, pp. 865–892, 2010.

[163] V. Henri and C. Henri, "On our earliest recollections of childhood," *Psychological Review*, vol. 2, pp. 215-216, 1895.

[164] S. Freud, "Childhood and concealing memories," in *The Basic Writings of Sigmund Freud*, The Modern Library, NY, USA, 1905/1953.

[165] P. J. Bauer and M. Larkina, "The onset of childhood amnesia in childhood: A prospective investigation of the course and determinants of forgetting of early-life events," *Memory*, vol. 22, no. 8, pp. 907–924, 2014.

[166] M. S. Seidenberg and L. M. Gonnerman, "Explaining derivational morphology as the convergence of codes," *Trends in Cognitive Sciences*, vol. 4, no. 9, pp. 353–361, 2000.

[167] M. Coltheart, K. Rastle, C. Perry, R. Langdon, and J. Ziegler, "DRC: a dual route cascaded model of visual word recognition and reading aloud," *Psychological Review*, vol. 108, no. 1, pp. 204–256, 2001.

[168] M. V. Butz and E. F. Kutter, *How the Mind Comes into Being: Introducing Cognitive Science from a Functional and Computational Perspective*, Oxford University Press, 2016.

[169] M. W. Harm and M. S. Seidenberg, "Phonology, reading acquisition, and dyslexia: Insights from connectionist models," *Psychological Review*, vol. 106, no. 3, pp. 491–528, 1999.

[170] L. M. Gonnerman, M. S. Seidenberg, and E. S. Andersen, "Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology," *Journal of Experimental Psychology: General*, vol. 136, no. 2, pp. 323–345, 2007.

[171] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[172] C. Clarke and P. Luce, "Perceptual adaptation to speaker characteristics: Vot boundaries in stop voicing categorization," in *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*, 2005.

[173] D. Norris, J. M. McQueen, and A. Cutler, "Perceptual learning in speech," *Cognitive Psychology*, vol. 47, no. 2, pp. 204–238, 2003.

[174] A. Schweitzer and N. Lewandowski, "Social factors in convergence of f1 and f2 in spontaneous speech," in *Proceedings of the 10th International Seminar on Speech Production*, Cologne, Germany, 2014.

[175] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behavioral and Brain Sciences*, vol. 27, no. 2, pp. 169–190, 2004.

[176] C. M. Clarke and M. F. Garrett, "Rapid adaptation to foreign-accented English," *The Journal of the Acoustical Society of America*, vol. 116, no. 6, pp. 3647–3658, 2004.

[177] W. J. Levelt, "Die konnektionistische mode," *Sprache Und Kognition*, vol. 10, no. 2, pp. 61–72, 1991.

[178] A. Kostić, "Informational load constraints on processing inflected morphology," in *Morphological Aspects of Language Processing*, L. B. Feldman, Ed., Lawrence Erlbaum Inc. Publishers, New Jersey, USA, 1995.

[179] J. P. Blevins, *Word and Paradigm Morphology*, Oxford University Press, 2016.

[180] E. Förstemann, *Über Deutsche volksetymologie. Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen*, 1852.

[181] K. Nault, *Morphological Therapy Protocal [Ph. D. thesis]*, University of Alberta, Edmonton, 2010.

Hindawi
Complexity
Volume 2018, Article ID 2674523, 15 pages
https://doi.org/10.1155/2018/2674523

Research Article

Spread the Joy: How High and Low Bias for Happy Facial Emotions Translate into Different Daily Life Affect Dynamics

Charlotte Vrijen,[1] **Catharina A. Hartman**,[1] **Eeske van Roekel**,[1,2] **Peter de Jonge**,[1,3] **and Albertine J. Oldehinkel**[1]

[1]*Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen and University Medical Center Groningen, Groningen, Netherlands*
[2]*Department of Developmental Psychology, Tilburg University, Tilburg, Netherlands*
[3]*Department of Developmental Psychology, University of Groningen, Groningen, Netherlands*

Correspondence should be addressed to Charlotte Vrijen; c.vrijen@rug.nl

Received 13 February 2018; Revised 29 June 2018; Accepted 12 August 2018; Published 2 December 2018

Academic Editor: Cynthia Siew

There is evidence that people commonly show a bias toward happy facial emotions during laboratory tasks, that is, they identify other people's happy facial emotions faster than other people's negative facial emotions. However, not everybody shows this bias. Individuals with a vulnerability for depression, for example, show a low happy bias compared to healthy controls. The main aim of this study was to acquire a better understanding of laboratory measures of happy bias by studying how these translate to people's daily life. We investigated whether stable high and low happy bias during a laboratory task were associated with different daily life affect dynamics (i.e., effects from one time interval of 6 hours to the next). We compared the daily life affect dynamics of young adults (age 18–24) with a high bias toward happy facial emotions ($N = 25$) to the affect dynamics of young adults with a low bias toward happy emotions ($N = 25$). Affect and related measures were assessed three times per day during 30 days. We used multilevel vector autoregressive (VAR) modelling to estimate lag 1 affect networks for the high and low happy bias groups and used permutation tests to compare the two groups. Compared to their peers with a low happy bias, individuals with a high happy bias more strongly sustained the effects of daily life reward experiences over time. Individuals with a high happy bias may use their reward experiences more optimally in daily life to build resources that promote well-being and mental health. Low reward responsiveness in daily life may be key to why individuals who show a low happy bias during laboratory tasks are vulnerable for depression. This study illustrates the potential benefits of a network approach for unraveling psychological mechanisms.

1. Introduction

It is an interesting phenomenon that some people have a tendency to be relatively fast in identifying other people's happy facial emotions while others are relatively fast in identifying other people's negative facial emotions. Happy bias is an implicit bias of which people are probably unaware themselves; therefore, it is commonly assessed with standardized laboratory tasks. There is consistent evidence from studies using these laboratory tasks that people generally show a bias toward happy facial emotions, that is, they commonly identify other people's happy facial emotions faster than other people's negative facial emotions [1]. However, not everybody shows this bias. Depressed individuals, for example, show a low happy bias compared to healthy controls [2–4], and there are indications that a low happy bias is already present before onset of depression and predicts onset of depression [5, 6]. Given these findings and the accumulating evidence that the smallest building blocks of an individual's adaptive and maladaptive affect patterns are found in daily life affect dynamics [7–9], one would expect that high and low happy bias also reflect differences in daily life affect dynamics. However, to date this has not been investigated. In the present study, we looked at daily life correlates of laboratory measures of happy bias. We investigated how happy bias during a standardized laboratory task translates to daily

life affective dynamics by comparing the daily life affect dynamics (i.e., effects from one time interval of 6 hours to the next) between young adults with a stable high happy bias and young adults with a stable low happy bias. The main aim of this study was to acquire a better understanding of the importance and scope of laboratory measures of happy bias in people's daily life. Our findings are expected to facilitate the interpretation of laboratory measures of happy bias and, because of our focus on adaptive and maladaptive affect dynamics, may possibly also provide clues to why a low happy bias is associated with an increased risk for depression.

Indications of which daily life affect dynamics promote mental health (i.e., are adaptive) and which ones are associated with depressive problems (i.e., are maladaptive) can be found in both laboratory studies and in studies based on ecological momentary assessments (EMA). Evidence from laboratory tasks suggests that the inability to sustain positive emotions [10, 11], the inability to sustain activation in neural circuits underlying positive affect and reward over time [12], and the incapability to disengage from negative self-referential rumination [13] are associated with depressive symptoms and clinical depression. It was further found that positive affect facilitates recovery from negative emotional experiences [14, 15]. EMA studies also indicate that the inability to sustain positive affect over time in daily life is associated with depressive symptoms (e.g., anhedonia), in general as well as in clinically depressed populations [16, 17]. Additionally, the ability to generate positive affect from pleasant experiences in daily life predicted fewer symptoms of depression and anxiety in individuals with a history of depression [18] and in individuals who had been exposed to childhood adversity or recent stressful life events [19]. Taken together, this evidence suggests that the following types of affect dynamics are adaptive and promote mental health:

(1) The ability to sustain positive affect and positive experiences over time [10–12, 16, 17], that is, the carry-over of positive affect and positive experiences from one time interval to the next

(2) The ability to use positive experiences to generate positive affect and vice versa [18, 19], that is, the carry-over of positive experiences to positive affect and vice versa from one time interval to the next

(3) The ability to use positive affect and positive experiences to dampen negative affect, negative thoughts (i.e., rumination), and negative experiences [13–15].

In the present study, we investigated whether a high happy bias as compared to a low happy bias during a laboratory task is associated with (1) enhanced responses to positive affect and positive experiences over time in daily life, with more carry-over over time (i.e., from one 6-hour time interval to the next), and more carry-over from one type of positive affect or positive experience to another, and a stronger dampening effect on negative affect, thoughts, and experiences; and (2) diminished responses to negative affect, thoughts, and experiences in daily life, with less carry-over over time (i.e., from one time interval of 6 hours to the next), less carry-over from one type of negative affect, thoughts, or experience to another, and weaker dampening effects on positive affect and positive experiences.

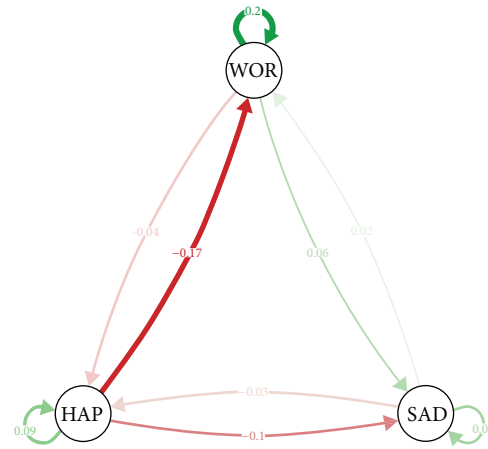


Figure 1: Fictitious example of a temporal network containing three nodes. The green edges represent positive directed associations; for example, on average high levels of worrying during one assessment predict high levels of sadness during the next assessment. The red edges represent negative directed associations; for example, on average high levels of happiness during one assessment predict low levels of worrying during the next assessment. The self-loops represent autocorrelations (i.e., the effect of the node on itself from one assessment to the next).

We used a network approach to affect dynamics, which entails that psychological symptoms or constructs are represented as interacting components of a complex dynamic system [20, 21] and that these dynamics define the very nature of the psychological phenomena we study (i.e., mental disorders, well-being) [22]. This approach can be used to investigate cross-sectional associations between symptoms at a specific point in time, but also, as in the present study, to investigate the temporal dynamics of affect over time. These temporal networks consist of "nodes" (i.e., the variables in the network) and "edges" (i.e., the directed associations between these nodes from one assessment to the next). See Figure 1 for a fictitious example of a temporal network containing the three nodes "happiness" (HAP), "sadness" (SAD), and "worrying" (WOR). In this network model, as well as in the models we used, each node is predicted by the lag (i.e., $t-1$) of all other variables and itself. SAD at time t is, for example, predicted by HAP_{t-1}, WOR_{t-1}, and SAD_{t-1}. Temporal networks can be used to study how different affect components interact as a dynamic system over time. They provide insightful visualizations of the interplay of the network components, and it is also possible to compute centrality indices indicating the importance of each of the components in the network.

We compared the daily life dynamic affect networks (i.e., effects from one time interval of 6 hours to the next) of two groups of young adults with extreme and stable biases to happy facial emotions during a laboratory task. We selected a high happy bias group, consisting of individuals who were considerably more sensitive to happy emotions than to

negative emotions, and a low happy bias group, who showed considerably less bias toward happy emotions, or even a bias toward negative emotions.

The affect dynamics of the high and low happy bias groups were compared on nodes that are associated with reward responsiveness, emotion regulation, and depressive symptoms. We selected three nodes that were related to positive affect and positive experiences (for the sake of brevity and readability hereafter referred to as positive nodes): "feeling joyful," "pleasant experiences," and "feeling interested"; and four nodes related to negative affect, negative thoughts, and negative experiences (for the sake of brevity and readability hereafter referred to as negative nodes): "feeling sad," "feeling irritated," "worrying," and "unpleasant experiences." The nodes feeling interested, feeling sad, and feeling irritated closely resemble core symptoms of depression according to the Diagnostic and Statistical Manual of Mental Disorders [23], and feeling interested also reflects openness to new experiences and an inclination to actively approach and explore the outside world. The nodes feeling joyful and pleasant experiences are particularly relevant in the light of indications that high transference of positive emotions over time in daily life [17] and the ability to generate boosts of positive affect from pleasant daily life experiences [19] may protect against affective problems. As opposed to feeling interested, feelings of joy and pleasant experiences are by definition rewarding at the very moment they are experienced. To illustrate the difference, people may cheer because they feel joy or pleasure, but not because they feel interested. More than the other nodes, "pleasant experiences" and "unpleasant experiences" reflect not only affective states but also the type of events individuals are involved in and their ability to seek out rewarding experiences or escape from a cascade of negative events. The term "worrying" was used for both negative thoughts about the past, often referred to as "rumination", and negative thoughts about the future, commonly referred to as "worrying" [24, 25]. "Worrying" was included for its associations with depressive disorder [26], positive and negative affect [24], and reduced cognitive control [27]. It was explored how these different nodes interact as dynamic systems and if these dynamics differed between the high and the low happy bias group.

We expected that an increase in positive affect and positive experiences, particularly of the directly rewarding positive nodes joy and pleasant experiences, would have larger and longer-lasting effects in the high happy bias group than in the low happy bias group. More specifically, we expected that the high happy bias group would more easily sustain and act upon pleasant experiences and feelings of joy to enhance positive affect and positive experiences and dampen negative affect, negative thoughts, and negative experiences. We also expected that pleasant experiences would generalize or carry over to feelings of joy and the other way around. We thus hypothesized that pleasant experiences and feelings of joy would be stronger predictors in the network of the high happy bias group than in the network of the low happy bias group (hypothesis 1) and that the nodes pleasant experiences and joy would more strongly predict themselves (i.e., pleasant experiences and feelings of joy would be more easily sustained) and each other (i.e., more carry-over between pleasant experiences and feelings of joy) over time in the high happy bias group than in the low happy bias group (hypothesis 2) and that joy and pleasant experiences would more strongly predict the negative affect nodes over time (i.e., a larger dampening effect on negative nodes) in the high happy bias group than in the low happy bias group (hypothesis 3). Further, because of the hypothesized reduced reward responsiveness in the low happy bias group, we expected the negative affect nodes to be stronger predictors in the network of the low happy bias group than in the network of the high happy bias group (hypothesis 4) and that the negative affect nodes would more strongly predict themselves and each other over time in the low happy bias group than in the high happy bias group (hypothesis 5) and more pronounced negative associations between negative nodes and positive nodes over time (i.e., a larger dampening effect on positive nodes) in the low happy bias group than in the high happy bias group (hypothesis 6). Although feeling interested is a positive node, we did not expect it to have a similar role as joy and pleasant experiences as we consider feelings of joy and pleasure as intrinsically rewarding, whereas feeling interested is a more instrumental node, which only potentially leads to reward. More specifically, rather than group differences in the way in which feeling interested influenced other nodes, we expected that joy and pleasant experiences would more strongly predict interest in the high than in the low happy bias group (hypothesis 7).

2. Methods

2.1. Sample. Data were collected as part of the "No Fun No Glory" (NFNG) study, in which we investigated anhedonia in young adults. The study was approved by the Medical Ethical Committee from the University Medical Center Groningen (no. 2014/508) and registered in the Dutch Clinical Trial Register (NTR5498). Participants were treated in accordance with the Declaration of Helsinki and indicated their informed consent prior to enrollment in the study. The project started with a large online screening survey in the northern part of the Netherlands among 2937 young adults between 18 and 24 years old. Participants were recruited through advertisements on electronic learning environments of university and higher and intermediate vocational education institutes, pitches during lectures and classes, flyers, and advertisements on social media. After subscribing on the study website (http://www.nofunnoglory.nl), participants received an email with the link to the online survey. The survey contained questionnaires about, for example, pleasure, psychiatric problems, and stress, as well as a facial emotion identification task. From the screening survey, 69 young adults who suffered from persistent anhedonia and 69 controls were selected for the part of the study in which momentary assessments were completed. For a description of the selection procedure for the anhedonia and control group in the NFNG project, see Section 1 of the online Supplementary Material or van Roekel et al. [28]. From the 138 participants who completed the momentary assessments during the first month, we selected 25 participants with a high happy

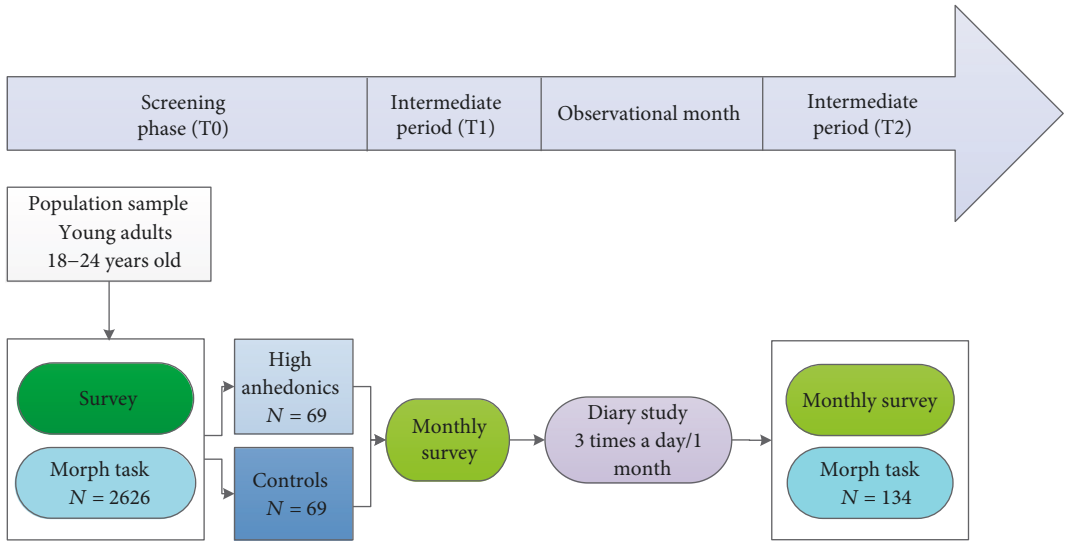


Figure 2: Flowchart of morph tasks and diary study month (i.e., momentary assessments) used in the current study.

bias and 25 participants with a low happy bias for the present study.

2.1.1. Selection of High and Low Happy Bias Groups. We used extreme bias groups rather than the full happy bias continuum because of both conceptual and methodological considerations. First and foremost, the extreme-group approach more closely fitted our supposition that mainly happy biases in the extremes of the distribution distinguish between adaptive and maladaptive affective mechanisms [29]. Second, a particular strength of network analyses is that these can be used to explore group differences in overall patterns of affect dynamics rather than investigating single effects only. Estimating and plotting the networks for each of the groups separately yields more insight into the affect dynamics within these groups than a single network based on the total sample, while it is still possible to test statistically whether specific affect patterns differ between the groups.

We selected participants for the high and low happy bias groups without taking into account whether participants belonged to the anhedonia or the control group. The selection was based on scores on a facial emotion identification morph task participants completed for the first time as part of the online screening survey (T0) and a second time after the first month of momentary assessments (T2); see Figure 2 for a flowchart.

We excluded four participants who did not complete the morph task at T2. During the morph task, participants were shown 24 10-second movie clips of neutral faces which slowly changed into one of four emotions: happy, sad, angry, or fearful. The participants had to press the spacebar as soon as they identified the emotion the neutral face turned into. For a more detailed description of the morph task, which was a shortened version of a task developed at Radboud University Nijmegen, the Netherlands [30], see Section 1 of the Supplementary Material or Vrijen et al. [29]. For each participant, the mean reaction time (RT) of correctly identified trials was calculated per emotion, resulting in RT Happy, RT Sad, RT Angry, and RT Fearful. We excluded one participant with less than 50% correct answers at T2. Separate happy bias scores were calculated at T0 and T2 by dividing the average of RT Sad, RT Angry, and RT Fearful by RT Happy. A higher happy bias means being faster in identifying happy emotions than in sad, angry, and fearful emotions.

We were interested in the affect dynamics associated with trait high and low happy bias and compared the average affect dynamics during 30 days of individuals with a stable high happy bias (i.e., stable during these 30 days) to the affect dynamics of individuals with a stable low happy bias. Because stable happy bias and state fluctuations can only be unraveled by using happy bias at two time points, we selected an extreme high stable and an extreme low stable happy bias group based on the ranked happy bias scores at T0 and T2. Happy bias at T0 and Happy bias at T2 were each ranked from low (ranking 133) to high (ranking 1), and selection of the two happy bias groups was based on the summed ranks for T0 and T2. The 25 participants with the highest summed rank were selected for the high happy bias group, and the 25 participants with the lowest summed rank for the low happy bias group. An additional advantage of this approach was that part of the measurement error is also parceled out because a participant is only selected for the high happy bias group if scores on both tasks are high. (Please note that we do not mean to suggest that all differences between happy bias at T0 and T2 are due to measurement error. We acknowledge that there may well be state happy bias fluctuations within a person between T0 and T2, but in the present study, we are interested in the daily life affect dynamics associated with more stable high and low happy bias.) To ensure that high (or low) scores reflected a high (or low) score relative to the rest of the group on both tasks, we used summed rank scores rather than summed mean scores. In this way, scores on both tasks count equally even in the case of general learning effects of the whole group. The middle group, which consisted of participants with moderate or unstable happy bias scores, was excluded from the main analyses.

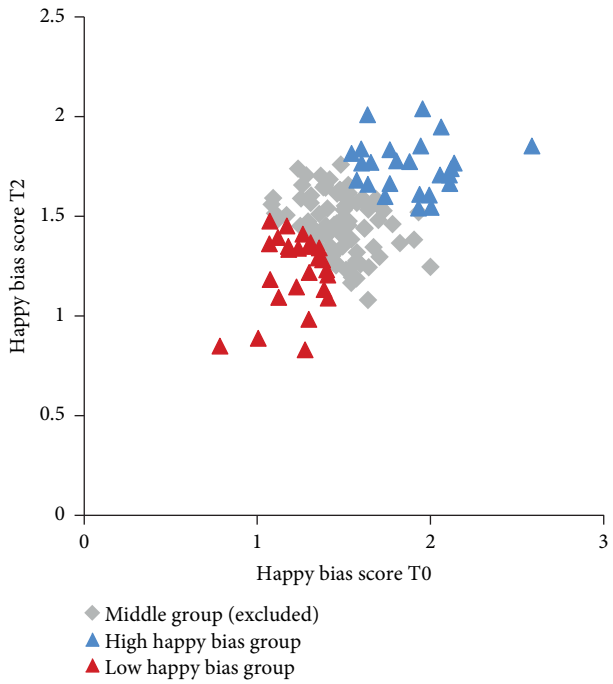


Figure 3: Happy bias scores T0 and T2 and selection high and low happy bias group.

This group was taken into account in part of the post hoc sensitivity checks (see Section 4 of the Supplementary Material). See Figure 3 for the individual happy bias scores at T0 and T2 for the high happy bias group, the low happy bias group, and the middle group.

2.2. Ecological Momentary Assessments. In the online questionnaire participants filled out three times a day, we included items to measure positive affect and positive experiences, negative affect, negative thoughts, and negative experiences, social company, activities, etc. See van Roekel et al. [28] for a detailed description of all momentary items that were assessed as part of the No Fun No Glory Study. Starting point as well as times of receiving the questionnaires during the day were personalized to the schedule and preference of the participants. They received a text message on their smartphone with the link to the questionnaire three times a day on fixed times with 6-hour intervals in between (e.g., 10:00 AM, 4:00 PM, and 10:00 PM). The questionnaire had to be completed within 2 hours after the first notification. If necessary, reminders were sent after 1 hour and again after 1.5 hours. Completion of the questionnaires took on average 3 minutes.

We included the following items in the present study: Since the last assessment I felt joyful (JOY); Think about the most pleasant event you experienced since the last assessment: how pleasant was this experience for you? (POS); Since the last assessment I felt interested in the things around me (INT); Since the last assessment I felt sad (SAD); Since the last assessment I felt irritated (IRR); Since the last assessment I have been worrying (WOR); Think about the most unpleasant event you experienced since the last assessment: how unpleasant was this experience for you? (NEG). Because we considered JOY, INT, SAD, and IRR to be sensitive to overnight recall bias, the morning assessments of these items were phrased more momentarily, that is, into "I feel joyful," "I feel interested in the things around me," "I feel sad," and "I feel irritated." Participants indicated their endorsement to these items by means of a slider on a Visual Analogue Scale (VAS), with "not at all" as its left anchor and "very much" as its right anchor. The position of the slider was transformed into a score between 0 ("not at all") and 100 ("very much").

2.3. Statistical Analyses. We provided descriptive statistics for gender, age, education, and anhedonia status; calculated mean levels of all variables used in this study; and showed them for the high and low happy bias group separately. We used R package psych version 1.6.12 [31, 32] to calculate the group and individual mean squared successive differences (MSSDs) for each node, as these indicate the amount of variability from one assessment to the next. We also indicated per node how many participants had an MSSD < 50, which was used as a criterion for insufficient variability from one assessment to the next, following Van Der Krieke et al. [33]. If one group shows a higher average MSSD than the other or contains more participants with an acceptable MSSD, it is possible that this results in more power for this group. Therefore, we decided that if more or stronger significant associations were found for this group, we would address the possibility that these differences were driven by differences in MSSD in the discussion.

We used multilevel vector autoregressive (VAR) modelling in R package mlVAR version 0.4 [34, 35] to explore the daily life dynamics between JOY, POS, INT, SAD, IRR, WOR, and NEG for the high and low happy bias groups. One of the main advantages of mlVAR was the availability of tools that, in combination with the R packages igraph [36] and qgraph [37], allowed not only visualization of networks and centrality indices on a group level but also visualization of individual variation within groups.

Although exact power calculations are not possible for VAR analyses, a minimum of 50 assessments per person has been recommended for individual VAR analyses [38]. We performed multilevel VAR analyses for which power is influenced by both the number of assessments and the number of persons. Our analyses were based on three assessments per day for a period of 30 days, that is, 90 assessments per person (with an average of 6 missings per person), and our high and low happy bias groups consisted of 25 participants each. We ensured sufficient power by limiting the number of parameters estimated in mlVAR; that is, we focused mainly on within-subject processes, refrained from investigating the influence of between-subject predictors, and did not estimate correlations between random effects (see below for further details). We have performed a simulation study based on the present study's effect sizes and number of subjects and data points. Eight hundred datasets were simulated in which the individual network models we found were generated as the "true" models. Fixed and random effects were estimated with the same method and number of nodes as in our main mlVAR analyses. In the present study, we used on average 84 time points per subject, and for this number

of time points the simulation study showed high correlations between the true and the estimated fixed and random effects of the simulated datasets (see Figure S1 in the Supplementary Material). This indicates that the method we used is appropriate for our effect sizes, number of subjects, and number of time points. Additionally, because all of our hypotheses were based on network patterns rather than on specific effects of a single variable, our findings do not rely on single paths in the network models.

As a first preparatory step, we removed linear time trends from the data, because time trends violate the stationarity assumption of VAR analyses and may bias parameter estimation [39]. We also removed cyclic time of day trends prior to VAR analysis, because mlVAR does not allow controlling for time of day. Linear and cyclic time trends were removed by regressing each variable on time and on dummy variables for afternoon and evening, within each individual. The residuals from these analyses were used as input for the VAR models.

For estimating networks containing both autoregressive and cross-lagged effects, it has been recommended to person-mean center all predictors prior to the analyses in order to separate within-subject from between-subject effects [40, 41]. Because our main interest was to grasp daily life psychological processes which take place within individuals, we separated within-subject from between-subject effects even further by within-person standardization of all network variables prior to the VAR analyses. For comparing the relative strengths of different predictors within and between networks, standardization of the coefficients has been recommended because differences in coefficients may be due to differences in variance [42, 43]. Using raw coefficients to compute centrality indices has been discouraged as well [42].

In mlVAR, separate lag 1 networks were estimated for the high and low happy bias groups, by means of the lmer function from the linear mixed-effects R package lme4 version 1.1-15 [44]. The networks were constructed by performing seven univariate multilevel VAR analyses, one for each dependent variable, and combining the results into a network. In each of the univariate multilevel VAR analyses, the dependent variable was predicted by the lag (i.e., $t-1$) of all other variables and itself. This means that, for example, feeling irritated (IRR) at time t was predicted by INT_{t-1}, JOY_{t-1}, SAD_{t-1}, WOR_{t-1}, POS_{t-1}, NEG_{t-1}, and IRR_{t-1}. The unique direct temporal effects were modeled [22, 42]. Random effects were estimated to account for individual differences. We assumed no correlations between random intercepts and random slopes (orthogonality specification in mlVAR), as the person-mean of each variable was equal to 0 after within-person standardization.

The above-described procedure resulted in a network for the high happy bias and a network for the low happy bias group. For each node of these networks, we calculated two centrality indices, outstrength and instrength. The outstrength of a node represents the summed strength, that is, the absolute value of the coefficients, of all outgoing paths from this node at $t-1$ to other nodes at time t, and as such reflects how strongly the node predicts other nodes over time. The instrength reflects how strongly a particular node is predicted by other nodes over time and is computed by the summed strength of all its incoming paths at time t from other nodes at $t-1$. In mlVAR, the packages igraph version 1.1.2 and qgraph version 1.4.4 were used to plot the networks and to compute and visualize the centrality indices. Autoregressive components were not included in the outstrength and instrength [37]. We compared the group network models and centrality indices of the high and low happy bias group by means of visual inspection. Next, we explored individual differences within the two groups by plotting the instrength and outstrength for each person separately, based on the person-specific effects.

In addition to the visual comparisons of the networks and centrality indices, we performed seven permutation tests to test the hypothesized differences between the high and low happy bias groups. Significant results on the permutation test suggest differences between high and low happy bias in the population. The permutation tests compared the observed differences of interest to distributions of possible differences under the null hypothesis of no differences between the groups. Distributions of possible differences were derived from reshuffling the groups randomly 10,000 times, also called Monte Carlo sampling. For each reshuffle, differences between the two reshuffled groups were estimated with the lmer function of R package lme4, that is, in the same way as the original models had been defined in mlVAR. If an observed difference between the high and low happy bias groups was within 2.5% on either side of the distribution of the 10,000 possible differences, the difference between the high and low happy bias group was considered significant (i.e., $p < 0.05$). We used an adapted version of the permutation test developed by Snippe et al. [45] to test differences between the high and low happy bias groups which match our hypotheses as described in the Introduction. See Table 1 for a description of the seven hypotheses and their operationalization for the permutation tests.

All of the tested difference scores were based on the fixed effects of the group models. For permutation tests (1) and (4), we used absolute edge weights in order to avoid that expected positive and expected negative associations cancel each other out. All of our hypotheses and therefore also all permutation tests applied to outstrength. We explored possible differences in instrength between the high and low happy bias groups by visual comparison of the networks and centrality plots and did not use permutation tests because we did not have clear hypotheses in advance.

Finally, we performed multiple sensitivity analyses to explore the robustness of our findings. First, we repeated the mlVAR analyses in Mplus version 8 [46], which allowed multivariate mlVAR analyses with a Bayesian estimator. Second, although the decision to use extreme groups rather than continuous happy bias measures was driven by valid conceptual and methodological considerations, there were no clear criteria on how extreme the groups should be and therefore the exact number of individuals selected for each group (i.e., 25) was somewhat arbitrary. To assess the robustness of the results based on groups of 25 individuals, we estimated the networks, computed the centrality indices, and performed the permutation tests for bias groups of 20, 30, 35,

Table 1: Description and operationalization of the seven hypotheses tested with the permutation tests.

	Description	Permutation test
Hypothesis 1	JOY and POS are stronger predictors in the network of the high happy bias group than in the network of the low happy bias group	(1) The total summed absolute edge weight of all outgoing edges from JOY and POS at time $t-1$ to all nodes in the network at time t (including autoregressive edges) is larger for the high happy bias group than for the low happy bias group
Hypothesis 2	JOY and POS more strongly predict themselves (i.e., are more easily sustained over time) and each other (i.e., more carry-over between JOY and POS) over time in the high happy bias group than in the low happy bias group	(2) The total summed edge weight of all outgoing edges from JOY and POS at time $t-1$ to JOY and POS at time t (including autoregressive edges) is larger for the high happy bias group than for the low happy bias group
Hypothesis 3	JOY and POS more strongly predict the negative nodes (i.e., larger dampening effect on negative nodes) over time in the high happy bias group than in the low happy bias group	(3) The total summed edge weight of all outgoing edges from JOY and POS at time $t-1$ to SAD, IRR, WOR, and NEG at time t is larger for the high happy bias group than for the low happy bias group
Hypothesis 4	The negative nodes are stronger predictors in the network of the low happy bias group than in the network of the high happy bias group	(4) The total summed absolute edge weight of all outgoing edges from SAD, IRR, WOR, and NEG at time $t-1$ to all nodes in the network at time t (including autoregressive edges) is larger for the low happy bias group than for the high happy bias group
Hypothesis 5	The negative nodes more strongly predict themselves (i.e., are more easily sustained over time) and each other (i.e., more carry-over between the negative nodes) over time in the low happy bias group than in the high happy bias group	(5) The total summed edge weight of all outgoing edges from SAD, IRR, WOR, and NEG at time $t-1$ to SAD, IRR, WOR, and NEG at time t (including autoregressive edges) is larger for the low happy bias group than for the high happy bias group
Hypothesis 6	More pronounced negative associations between negative nodes and JOY and POS (i.e., larger dampening effect on JOY and POS) over time in the low happy bias group than in the high happy bias group	(6) The total summed edge weight of all outgoing edges from SAD, IRR, WOR, and NEG at time $t-1$ to JOY and POS at time t is larger for the low happy bias group than for the high happy bias group
Hypothesis 7	JOY and POS more strongly predict INT in the high than in the low happy bias group	(7) The total summed edge weight of all outgoing edges from JOY and POS at time $t-1$ to INT at time t is larger for the high happy bias group than for the low happy bias group

JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in the things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences.

and 40 individuals. We used the same mlVAR methods as in the main analyses. As a third check, to adjust for anhedonia status, we computed subject-specific centrality indices based on the random estimates of the edges of the low and high happy bias networks and subsequently regressed the subject-specific centrality indices on anhedonia status and happy bias. See Sections 3–5 in the Supplementary Material for further details.

3. Results

3.1. Descriptive Statistics General Demographics. The high and low happy bias groups were quite comparable in terms of age, gender, and education (see Table 2). Although in the low happy bias group more participants attended university, in both groups all participants were enrolled in higher education. The groups differed considerably in symptoms of anhedonia.

Table 2: General demographics and anhedonia status.

	High happy bias group ($n = 25$) Mean (SD)/count (%)	Low happy bias group ($n = 25$) Mean (SD)/count (%)
Age	21.64 (1.77)	20.69 (2.05)
Females	20 (80%)	22 (88%)
University education	13 (52%)	18 (72%)
Higher vocational education	12 (48%)	7 (28%)
Anhedonic[a]	8 (32%)	13 (52%)
Control[a]	17 (68%)	10 (40%)
Switcher[a]	0 (0%)	2 (8%)

[a]Participants were classified as anhedonic or control if they met all criteria at T0 and did not change in pleasure levels from one group to the other at either T1 or T2. Otherwise, they were classified as switcher.

The descriptive statistics for the facial emotion identification variables are presented in Table 3. The high happy bias group had a mean happy bias score of 1.82, which means that happy facial emotions were identified on average 1.82 times faster than the negative facial emotions sadness, anxiety, and fear. The low happy bias group had a mean happy bias score of 1.23, which indicates that this group is on average

Table 3: Descriptive statistics of facial emotion identification scores.

	High happy bias group ($n = 25$)		Low happy bias group ($n = 25$)	
	Mean	SD	Mean	SD
RT Total	5522.61	574.66	5236.69	848.37
RT Happy	3449.08	377.55	4550.20	1034.25
RT Sad	6797.62	801.18	6051.65	1014.42
RT Angry	5761.44	705.29	5189.62	916.74
RT Fearful	6094.69	883.37	5168.86	821.34
Happy bias score	1.82	0.14	1.23	0.13
Happy bias rank	36.84	18.92	218.96	22.25

RT = reaction time; RT Total = mean score on RT Happy, RT Sad, RT Angry, and RT Fearful; happy bias score = mean score on RT Sad, RT Angry, and RT Fearful divided by RT Happy; happy bias rank = summed rank of happy bias score at T0 and T2.

Table 4: Descriptive statistics of the momentary assessment items used as nodes in the networks.

	Mean		Average within-person SD		Average within-person MSSD (n with MSSD < 50)	
	High bias $n = 2094$	Low bias $n = 2095$	High bias	Low bias	High bias	Low bias
JOY	60.27	56.07	12.98	12.38	237 (0)	248 (0)
POS	63.35	60.06	14.76	14.02	302 (0)	312 (0)
INT	58.31	53.30	13.28	13.35	263 (0)	275 (1)
SAD	13.96	18.49	11.16	12.34	225 (5)	262 (3)
IRR	15.42	19.87	13.02	15.03	302 (2)	384 (1)
WOR	21.88	22.88	13.75	15.26	291 (0)	310 (2)
NEG	32.82	39.27	18.75	18.79	565 (1)	554 (0)

JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in the things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences; MSSD = average within-person mean squared successive difference. *Note*. These descriptive statistics are based on data from which linear time trends, and cyclic time of day trends have already been removed.

still faster in identifying happy facial emotions, but the difference between happy and the negative emotions is only small.

Table 4 presents the descriptive statistics of the network variables. On average, the high happy bias group scored higher than the low happy bias group on the positive nodes (JOY, POS, and INT) and lower on the negative nodes (SAD, IRR, WOR, and NEG). Within-person SDs were quite similar across the groups, with the largest differences for IRR and WOR. MSSDs of all nodes were larger in the low happy bias group than in the high happy bias group and in both groups for all nodes MSSD ≥ 50 for almost all participants; that is, in the high happy bias group 5 participants had an MSSD < 50 on SAD, 2 on IRR, and 1 on NEG, and in the low happy bias group, 3 participants had an MSSD < 50 on SAD, 1 on INT, 1 on IRR, and 2 on WOR. Both the high and low happy bias group showed low numbers of missings per person on the momentary assessments; for both groups, the mean number of missings per person was 6.2 (out of 90), with min = 1 and max = 17.

3.2. Descriptive Statistics Network Models. The network models for the high happy bias group and the low happy bias group are visualized in Figure 4; only the significant edges with p values < 0.05 are depicted (see Table S1 in the Supplementary Material for the exact coefficients and significance levels of all paths). Green edges represent positive, and red edges represent negative associations from one node at time $t - 1$ to another node at time t; the thickness of the edges indicates the strength of the associations. As we within-person standardized all variables, the edge coefficients represent the change in terms of within-person SD in the outcome variable based on one within-person SD increase in the predictor variable.

3.2.1. Associations between Positive Nodes at Time t − 1 and Positive Nodes at Time t. For both groups, all autocorrelations of the positive nodes JOY, POS, and INT were significant. Autocorrelations of JOY and POS were higher, and JOY, POS, and INT were more densely connected to each other in the high happy bias group than in the low happy bias group. The positive edges suggest that an increase in one of the positive nodes is associated with an increase in the others at the next measurement.

3.2.2. Associations between Negative Nodes at Time t − 1 and Negative Nodes at Time t. For both groups, we found significant autocorrelations of the negative nodes WOR, NEG, and SAD, with a higher autocorrelation for WOR in the low happy bias group than in the high happy bias group. The autocorrelation of IRR was only significant in the low happy bias group. The negative nodes SAD, IRR, WOR, and NEG showed several positive temporal interrelations in both groups.

3.2.3. Associations between Positive Nodes at Time t − 1 and Negative Nodes at Time t. For the high happy bias group, a higher score on POS predicted a lower score on WOR, and a higher score on JOY predicted lower scores on IRR and NEG. For the low happy bias group, we did not find negative edges from POS and JOY to negative nodes at the next measurement.

3.2.4. Associations Between Negative Nodes at Time t − 1 and Positive Nodes at Time t. A higher score on negative nodes was significantly associated with a lower score on positive nodes at the next measurement for the low happy bias group only; that is, WOR and IRR showed negative associations with JOY and INT.

3.3. Descriptive Statistics Centrality Indices

3.3.1. Centrality Plots on Group Level. Centrality plots for outstrength and instrength are presented in Figure 5(a). JOY and POS had the highest outstrength in the high happy bias group and the lowest outstrength in the low happy bias group, indicating that JOY and POS most strongly predicted the other nodes in the high happy bias group and least strongly predicted the other nodes in the low happy bias

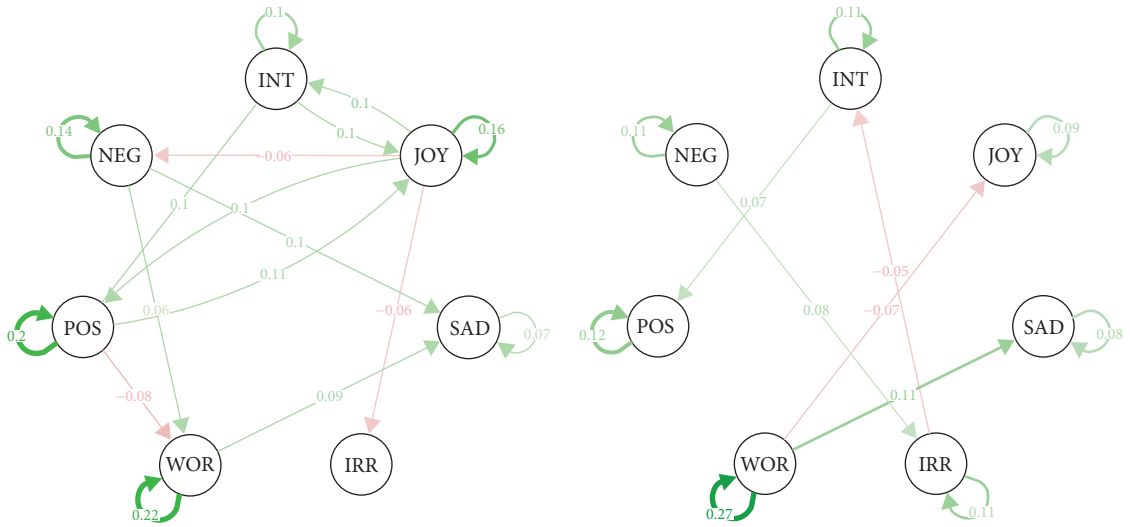


Figure 4: Significant association networks high happy bias group (a) and low happy bias group (b). JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences.

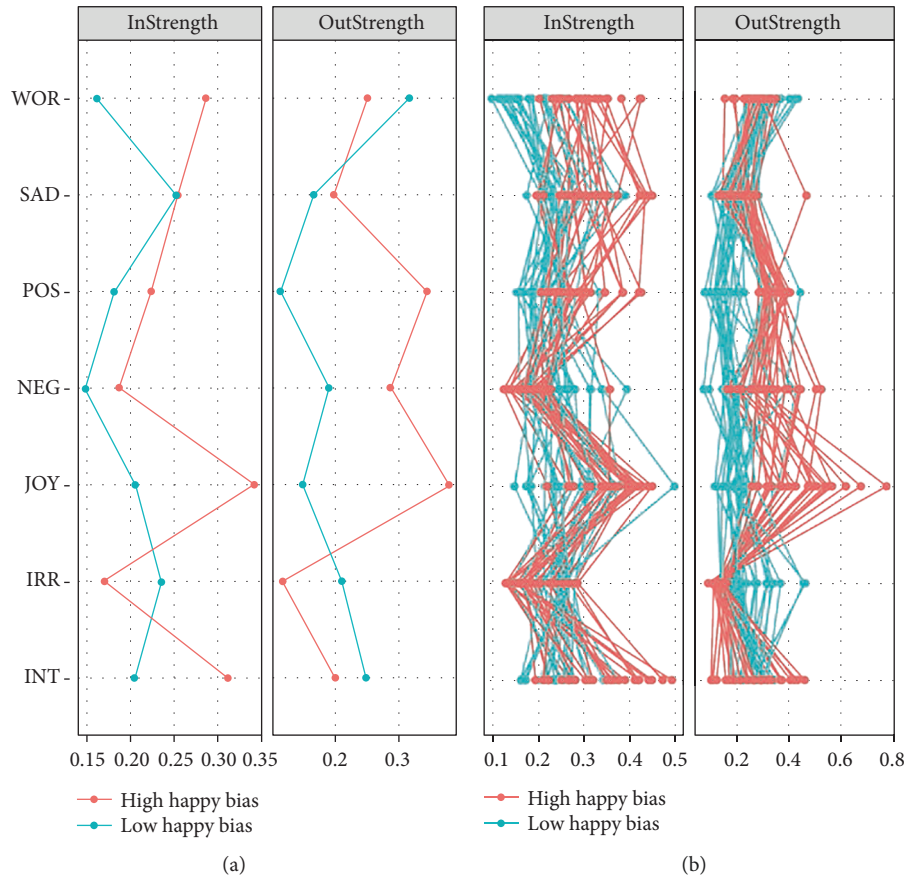


Figure 5: Centrality indices instrength and outstrength based on the complete network models. In panel (a) the indices are plotted for the high and low happy bias groups, and panel (b) illustrates the individual variation within these groups. JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences. *Note.* Estimations for panel (b) are based on multilevel group models (fixed effects) and individual variation within these groups (random effects). No separate individual models were estimated, and the lines in panel (b) should not be interpreted as such.

group. We found the largest differences in instrength between the high and low happy bias group for WOR, JOY, and INT, all of which were more strongly predicted by other nodes in the high happy bias group than in the low happy bias group.

3.3.2. Individual Variation in Centrality Indices. Individual variation in outstrength and instrength within the two happy bias groups is presented in Figure 5(b). In general, instrength showed more individual variation than outstrength. Regardless of the individual variation, the ranges of the high and low happy bias groups hardly overlapped on the outstrength of JOY, indicating that the low and high happy bias groups could be most clearly discriminated on the outstrength of this node.

3.4. Permutation Tests of Differences between the Low and High Happy Bias Group. As a brief reminder, the following seven hypotheses were tested: (1) JOY and POS are stronger predictors in the network of the high happy bias group than in the network of the low happy bias group; (2) JOY and POS more strongly predict themselves and each other over time in the high happy bias group than in the low happy bias group; (3) JOY and POS more strongly predict the negative nodes over time in the high happy bias group than in the low happy bias group; (4) the negative nodes are stronger predictors in the network of the low happy bias group than in the network of the high happy bias group; (5) the negative nodes more strongly predict themselves and each other over time in the low happy bias group than in the high happy bias group; (6) more pronounced negative associations between negative nodes and positive nodes over time in the low happy bias group than in the high happy bias group; and (7) JOY and POS more strongly predict INT in the high than in the low happy bias group. See Table 1 for a more detailed description of the hypotheses and their operationalization for the permutation tests.

Only permutation tests 1 and 2 reached significance at $p < 0.05$. The observed difference of the total absolute strength of all outgoing edges from JOY and POS between the high and low happy bias groups was 0.61, $p < 0.01$ (permutation test 1). The observed difference between the two groups of the total strength of all outgoing edges from JOY and POS to JOY and POS was 0.31, $p < 0.01$ (permutation test 2). We found no significant differences between the two groups for the total strength of all outgoing edges from JOY and POS to negative nodes (observed difference = −0.24, $p = 0.13$; permutation test 3), the total absolute strength of all outgoing edges of the negative nodes (observed difference = −0.14, $p = 0.57$; permutation test 4), the total strength of all outgoing edges from negative nodes to negative nodes (observed difference = −0.15, $p = 0.36$; permutation test 5), the total strength of all outgoing edges from negative nodes to JOY and POS (observed difference = 0.14, $p = 0.18$; permutation test 6), and the total strength of all outgoing edges from JOY and POS to INT (observed difference = 0.12, $p = 0.06$; permutation test 7).

3.5. Sensitivity Analyses. The multivariate Mplus multilevel VAR analyses showed minor differences for several of the network model estimates, but general patterns and main findings were confirmed (for more details, see Section 3 of the Supplementary Material). Networks, centrality indices, and permutation tests for extreme happy bias groups of 20, 30, 35, and 40 individuals showed the same patterns as the ones based on the original groups of 25 individuals. As expected, group differences became less pronounced as groups became larger and for $N = 40$ one of the two permutation tests was no longer significant. For more details, see Section 4 of the Supplementary Material. Finally, controlling for anhedonia status did not change our findings (see section 5 of the Supplementary Material for more details).

4. Discussion

Our study is the first to investigate what a bias for happy facial emotions as assessed by a standardized laboratory task pertains to in daily life. We found that feelings of joy and pleasant experiences were stronger predictors in the network of the high happy bias group than in the network of the low happy bias group (hypothesis 1) and that in the high happy bias group joy and pleasant experiences more strongly predicted themselves and each other over time (hypothesis 2). These were robust findings based on both visual inspection of the networks and centrality indices and permutation tests. Other group differences were only found by visual comparison but could not be corroborated by the permutation tests: joy and pleasant experiences dampened the negative nodes (i.e., sadness, irritation, worrying, and unpleasant experiences) in the high but not in the low happy bias network (hypothesis 3); the negative nodes dampened joy and pleasant experiences in the low but not in the high happy bias network (hypothesis 6); and joy and pleasant experiences predicted interest in the high but not in the low happy bias network (hypothesis 7). These group differences were present in the study sample, but since the permutation tests were not significant, it is not possible to draw inferences about group differences in the population. The fact that the permutation tests did not reveal group differences regarding these hypotheses may be due to large individual differences or small effects. We found no support that negative nodes more strongly predicted the overall affect network or the negative network in the low happy bias group than in the high happy bias group (hypotheses 4 and 5). Although the high and low happy bias groups differed considerably in symptoms of anhedonia, the differences we found between the high and low happy bias groups seem to be driven primarily by happy bias status and could not be (fully) explained by anhedonia status.

Our more specific results may be discussed in terms of the extent to which a certain node predicts other nodes at the next time point (outstrength) or in terms of the extent to which a certain node is predicted by other nodes at the previous time point (instrength). Because our hypotheses applied to outstrength and we did not have clear hypotheses about instrength in advance, our methodological approaches toward outstrength and instrength differed. For outstrength, we focused more on hypothesis-testing by means of

permutation tests, whereas we used a more exploratory approach based on visual inspection for instrength. Both perspectives will be discussed, starting with the outstrength.

We found that joy and pleasant experiences more strongly predicted affect over time in the high happy bias group than in the low happy bias group. This suggests that individuals with a high happy bias show a bias toward positive affect and positive experiences in their daily lives and may be better capable of sustaining positive affect and positive experiences over a longer period of time and generalizing it to other positive components than individuals with a low happy bias. This is particularly important because these same mechanisms, that is, the inability to sustain positive affect over time [10–12, 16, 17] and the inability to generate positive effect from pleasant experiences [18, 19], have been associated with depression in previous studies. In the present study, we also found indications that the same specific daily life affect dynamics that are associated with a low happy bias are also associated with depressive symptoms. That is, we found that for individuals suffering from anhedonia, which is one of the two core symptoms of depression, joy and pleasant experiences were weaker predictors of affect in the next six hours (see Section 5 of the Supplementary Material). Furthermore, daily life momentary positive affect during one month has been found to predict life satisfaction and a higher ability to adapt to changing environments after this month [47], which suggests that positive affect in the moment broadens one's attentional scope and facilitates building valuable cognitive and social resources essential to well-being [47, 48]. If feelings of joy can be sustained longer and spread to other positive experiences, their beneficial influence may be prolonged too. The savoring of positive affect, which includes the anticipation as well as the prolongation of positive affect, has been found to be associated with more life satisfaction and happiness and with lower levels of neuroticism, depression, and anhedonia [49]. In previous studies, it has also been found that positive affect facilitates recovery from negative experiences [14] and that resilient individuals use positive affect to downregulate negative affect [15]. This is in accordance with our findings that joy and pleasant experiences dampened negative nodes in the high but not in the low happy bias network and suggests that individuals with a high happy bias may be better equipped to use positive experiences and positive affect to regulate negative affect, thoughts, and experiences than the low happy bias group. However, caution is warranted in interpreting this finding because, although in all different sensitivity analyses joy and pleasant experiences dampened negative nodes in the high but not in the low happy bias group, the permutation test did not reach statistical significance. Therefore no conclusions can be drawn about group differences in the population. It is possible that the permutation test was not significant because of large individual variation in edges from joy and pleasant experiences to negative nodes, but this is only speculation. The regulation of negative nodes by means of positive nodes may also essentially occur on a shorter time frame, for example, 2 hours. If this is indeed the case, then the current method only picked up what was still left of the initial effect several hours later.

With regard to instrength, that is, the extent to which a certain node is predicted by other nodes at the previous time point, we found the largest differences between the high and low happy bias group for worrying, joyfulness, and feeling interested, all of which were more strongly predicted by other affect components in the high happy bias group than in the low happy bias group. A possible explanation is that this reflects psychological flexibility, such that for individuals with a high happy bias, worrying, joyfulness, and feeling interested are more dependent on context. How this might work can be illustrated by comparing the high and low happy bias networks in Figure 4. The level of worrying is influenced by pleasant and unpleasant experiences in individuals with a high happy bias, whereas for individuals with a low happy bias worrying seems to be less dependent on context, has a higher autocorrelation, and consequently tends to lead its own life. As psychological flexibility has been found to be highly important for optimal functioning in many situations, and psychological rigidity has been associated with depression as well as other forms of psychopathology [50, 51], the high happy bias group seems to be better off.

Sensitivity analyses were performed to explore the effects of different estimators, statistical packages, group sizes, a multivariate approach, and controlling for anhedonia status. All of these sensitivity analyses but one supported the original main findings completely; for the largest happy bias groups ($N = 40$), only partial support was found in the sense that one of the two original main findings was no longer significant. As expected, group differences became less pronounced as groups became larger. Two plausible explanations are, first, that for the happy bias groups of $N = 40$ the stability assumption of mlVAR analysis was not met, and second, that only happy bias in the extremes of the distribution may be associated with the development of adaptive versus maladaptive affective patterns in daily life.

Strengths of our study are, first of all, that we combined the best of two worlds by using a multilevel approach in which within-subject effects were separated from between-subject effects by within-person standardization of all variables prior to the analyses. This enabled us to explore dynamic processes that take place within individuals; at the same time, it allowed us to compare the two happy bias groups [43]. Secondly, following recent developments in the field [45, 52], in addition to visual comparison of the affect networks and centrality indices, we used permutation methods adapted to our specific hypotheses to test statistically whether the happy bias groups differed in their affect dynamics. A third strength of this study is its high ecological validity, as we assessed affect and related measures three times a day in daily life situations, for a period as long as 30 days, and achieved compliance rates of at least 80%. Additionally, the use of a morph task allowed us to assess the identification of more subtle traces of emotions, which is assumed to give a more ecologically valid perspective than static full-intensity facial emotion identification tasks, as in daily life static full-intensity facial emotions are quite rare. Finally, we repeated the facial emotion identification task and based our selection of the happy bias groups on individuals' scores on both tasks. This enabled us to select only those

participants with a stable happy bias. This was necessary because the daily life affect networks were estimated over a period of 30 days and participants showing large shifts in happy bias from one happy bias group to the other during this period would have added noise to the network models.

Evidently there are also limitations to our study. First, our sample largely consisted of higher educated females, which may limit the generalizability of our findings because gender and level of education may moderate the associations we investigated [53–58]. Second, the selection of extreme and stable happy bias groups resulted in small groups of 25 participants, which limits generalizability and resulted in insufficient power to correct for anhedonia status or use multivariate multilevel analysis, which would require the estimation of additional parameters. The disadvantage of the univariate approach is that correlations between the dependent variables and between random effects of the dependent variables were not taken into account. We presented sensitivity analyses to show the effects of a multivariate approach, different group sizes, and controlling for anhedonia status. However, most of these alternative approaches required multiple conceptual and methodological concessions and can only be interpreted as proxies to our original models. Third, we offered a network approach in which only unique direct temporal effects were studied, and no shared effects [22, 42]. As such, our approach should be regarded as complementary to approaches that take into account shared variance. Fourth, our results were based on assessments that were on average six hours apart. We were unable to grasp dynamic processes that took place within a shorter time frame. Finally, a limitation of our study that applies to all nonexperimental study designs is that we cannot make inferences about true causality; our conclusions are confined to "Granger" causality, that is, if a variable at time $t-1$ contains unique information to predict a second variable at time t, it is said to Granger cause this second variable [59]. We investigated the directed associations between different affect components over time, and it is plausible that other factors that were not included in our models explain part of these dynamics and therefore no true causal claims can be inferred from our network models.

Further research is required, first of all to confirm our findings by replication in other samples. Because of the present study's small group sizes, the conclusions are tentative awaiting attempts to replicate. Second, the specific conditions in which happy bias influences daily life affect dynamics need to be explored, for example, how extreme the bias needs to be before predicting positive or negative outcomes regarding well-being or mental health. Third, although our findings are promising, it should be noted that the ability to sustain positive emotions has been operationalized in many different ways in previous studies and there are also inconsistencies and unresolved issues, for example with respect to autocorrelation. It has been found that a stronger daily life autocorrelation of positive emotions over time protects against depression [17] but also that strong autocorrelations, for positive as well as negative emotions, predict depression [51, 58]. Further research is needed to investigate adaptive and maladaptive effects of strong autocorrelations versus psychological flexibility. It seems plausible that strong autocorrelations may indicate resistance to change and thereby limit psychological flexibility, but equally plausible that no carry-over of positive affect and positive experiences over time (weak autocorrelation) may also not be very adaptive. It may be important to consider different time scales [60], to look at proportions of autocorrelation in relation to cross-lagged paths (relative influence of other nodes) and to distinguish between high autocorrelation with respect to low and high levels of positive and negative affect and experiences. Finally, depression is a heterogeneous construct and specific subtypes of depression may be differentially associated with affect dynamics. A low happy bias could reflect such a subtype, and our study suggests that it can be useful to take happy bias into account when studying affect dynamics. Depressed individuals with a low happy bias may show different affect dynamics compared to depressed individuals with a high happy bias, but this remains to be investigated.

5. Conclusions

We compared young adults with a high bias for happy facial emotions during a standardized laboratory task to peers with a low bias for happy facial emotions on their daily life affect dynamics, using a highly personalized approach in which we separated within-subject from between-subject effects. Our most important and robust finding was that joy and pleasant experiences more strongly predicted the affect network of the high happy bias group than that of the low happy bias group. These findings tentatively suggest that individuals with a high happy bias are more responsive to positive, rewarding, experiences, and emotions, and more easily sustain them, whereas positive experiences and emotions seem to be more short-lived in the daily life of individuals who lack this happy bias. We propose that high reward responsiveness may be reflected in both a high happy bias during facial emotion identification and the ability to sustain and generalize positive experiences and positive affect in daily life. This may be key to why individuals with a bias toward happy facial emotions are potentially more resilient to developing depression. By using a network approach to compare the daily life affect dynamics of individuals with a high and with a low happy bias, we came closer to understanding the daily life mechanisms behind high and low happy bias during a laboratory task. This novel perspective is valuable for interpreting facial emotion processing tasks, as are often assessed in clinical research and practice. The present study illustrates the potential benefits of a network approach for unraveling psychological mechanisms.

Data Availability

Data and syntax have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/4czv3/.

Disclosure

Preliminary results of the present study were presented in April 2017 at the biennial conference of the Society for Research in Child Development (SRCD; poster), in June 2017 at the biennial conference of the Society for Ambulatory Assessment (SAA; oral presentation), and in May 2018 at the 30th annual convention of the Association for Psychological Science (APS; poster).

Conflicts of Interest

The authors have no conflicts of interest.

Acknowledgments

Research reported in this publication was funded by a Vici grant (016.001/002) from the Netherlands Organization for Scientific Research, which was awarded to Prof. Dr. A.J. Oldehinkel. We are grateful to everyone who participated in our research. We thank Dr. S. Epskamp for his assistance regarding R package mlVAR and his willingness to implement options for within-person standardization, Dr. W. Viechtbauer for allowing us to adapt his R script for permutation tests for our own purposes, and Dr. L.F. Bringmann and Dr. E.H. Bos for the stimulating methodological discussions.

Supplementary Materials

Section 1: descriptions of the No Fun No Glory (NFNG) selection procedures and the facial emotion identification morph task. Section 2: the exact coefficients and significance levels of the main analyses, and the results of a simulation study which was used to assess the reliability of mlVAR for our specific sample size, number of time points, and model specifications. Sections 3–5: results of the sensitivity analyses performed to assess the robustness of our findings. *(Supplementary Materials)*

References

[1] J. M. Leppänen and J. K. Hietanen, "Positive facial expressions are recognized faster than negative facial expressions, but why?," *Psychological Research*, vol. 69, no. 1-2, pp. 22–29, 2004.

[2] C. Bourke, K. Douglas, and R. Porter, "Processing of facial emotion expression in major depression: a review," *Australian & New Zealand Journal of Psychiatry*, vol. 44, no. 8, pp. 681–696, 2010.

[3] J. Joormann and I. H. Gotlib, "Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia," *Journal of Abnormal Psychology*, vol. 115, no. 4, pp. 705–714, 2006.

[4] S. A. Surguladze, A. W. Young, C. Senior, G. Brébion, M. J. Travis, and M. L. Phillips, "Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression," *Neuropsychology*, vol. 18, no. 2, pp. 212–218, 2004.

[5] J. Joormann, L. Talbot, and I. H. Gotlib, "Biased processing of emotional information in girls at risk for depression," *Journal of Abnormal Psychology*, vol. 116, no. 1, pp. 135–143, 2007.

[6] C. Vrijen, C. A. Hartman, and A. J. Oldehinkel, "Slow identification of facial happiness in early adolescence predicts onset of depression during 8 years of follow-up," *European Child & Adolescent Psychiatry*, vol. 25, no. 11, pp. 1255–1266, 2016.

[7] M. L. Pe, K. Kircanski, R. J. Thompson et al., "Emotion-network density in major depressive disorder," *Clinical Psychological Science*, vol. 3, no. 2, pp. 292–300, 2015.

[8] T. J. Trull, S. P. Lane, P. Koval, and U. W. Ebner-Priemer, "Affective dynamics in psychopathology," *Emotion Review*, vol. 7, no. 4, pp. 355–361, 2015.

[9] J. T. W. Wigman, J. van Os, D. Borsboom et al., "Exploring the underlying structure of mental disorders: cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach," *Psychological Medicine*, vol. 45, no. 11, pp. 2375–2387, 2015.

[10] M. S. Horner, G. J. Siegle, R. M. Schwartz et al., "C'mon get happy: reduced magnitude and duration of response during a positive-affect induction in depression," *Depression and Anxiety*, vol. 31, no. 11, pp. 952–960, 2014.

[11] D. L. McMakin, C. D. Santiago, and S. R. Shirk, "The time course of positive and negative emotion in dysphoria," *The Journal of Positive Psychology*, vol. 4, no. 2, pp. 182–192, 2009.

[12] A. S. Heller, T. Johnstone, A. J. Shackman et al., "Reduced capacity to sustain positive emotion in major depression reflects diminished maintenance of fronto-striatal brain activation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 52, pp. 22445–22450, 2009.

[13] L. M. Hilt and S. D. Pollak, "Characterizing the ruminative process in young adolescents," *Journal of Clinical Child & Adolescent Psychology*, vol. 42, no. 4, pp. 519–530, 2013.

[14] B. L. Fredrickson and R. W. Levenson, "Positive emotions speed recovery from the cardiovascular sequelae of negative emotions," *Cognition and Emotion*, vol. 12, no. 2, pp. 191–220, 1998.

[15] M. M. Tugade and B. L. Fredrickson, "Resilient individuals use positive emotions to bounce back from negative emotional experiences," *Journal of Personality and Social Psychology*, vol. 86, no. 2, pp. 320–333, 2004.

[16] V. E. Heininga, E. Van Roekel, J. J. Ahles, A. J. Oldehinkel, and A. H. Mezulis, "Positive affective functioning in anhedonic individuals' daily life: anything but flat and blunted," *Journal of Affective Disorders*, vol. 218, pp. 437–445, 2017.

[17] P. Höhn, C. Menne-Lothmann, F. Peeters et al., "Moment-to-moment transfer of positive emotions in daily life predicts future course of depression in both general population and patient samples," *PLoS One*, vol. 8, no. 9, article e75655, 2013.

[18] M. Wichers, F. Peeters, N. Geschwind et al., "Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: a momentary assessment study," *Journal of Affective Disorders*, vol. 124, no. 1-2, pp. 191–195, 2010.

[19] N. Geschwind, F. Peeters, N. Jacobs et al., "Meeting risk with resilience: high daily life reward experience preserves mental health," *Acta Psychiatrica Scandinavica*, vol. 122, no. 2, pp. 129–138, 2010.

[20] D. Borsboom and A. O. J. Cramer, "Network analysis: an integrative approach to the structure of psychopathology," *Annual Review of Clinical Psychology*, vol. 9, no. 1, pp. 91–121, 2013.

[21] L. F. Bringmann, N. Vissers, M. Wichers et al., "A network approach to psychopathology: new insights into clinical longitudinal data," *PLoS One*, vol. 8, no. 4, article e60188, 2013.

[22] L. F. Bringmann, M. L. Pe, N. Vissers et al., "Assessing temporal emotion dynamics using networks," *Assessment*, vol. 23, no. 4, pp. 425–435, 2016.

[23] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, Arlington, VA, USA, 5th edition, 2013.

[24] K. A. McLaughlin, T. D. Borkovec, and N. J. Sibrava, "The effects of worry and rumination on affect states and cognitive activity," *Behavior Therapy*, vol. 38, no. 1, pp. 23–38, 2007.

[25] S. Nolen-Hoeksema, "Responses to depression and their effects on the duration of depressive episodes," *Journal of Abnormal Psychology*, vol. 100, no. 4, pp. 569–582, 1991.

[26] S. Nolen-Hoeksema, B. E. Wisco, and S. Lyubomirsky, "Rethinking rumination," *Perspectives on Psychological Science*, vol. 3, no. 5, pp. 400–424, 2008.

[27] M. Beckwé, N. Deroost, E. H. W. Koster, E. De Lissnyder, and R. De Raedt, "Worrying and rumination are both associated with reduced cognitive control," *Psychological Research*, vol. 78, no. 5, pp. 651–660, 2014.

[28] E. van Roekel, M. Masselink, C. Vrijen et al., "Study protocol for a randomized controlled trial to explore the effects of personalized lifestyle advices and tandem skydives on pleasure in anhedonic young adults," *BMC Psychiatry*, vol. 16, no. 1, p. 182, 2016.

[29] C. Vrijen, C. A. Hartman, G. M. A. Lodder, M. Verhagen, P. de Jonge, and A. J. Oldehinkel, "Lower sensitivity to happy and angry facial emotions in young adults with psychiatric problems," *Frontiers in Psychology*, vol. 7, 2016.

[30] G. M. A. Lodder, R. H. J. Scholte, L. Goossens, R. C. M. E. Engels, and M. Verhagen, "Loneliness and the social monitoring system: emotion recognition and eye gaze in a real-life conversation," *British Journal of Psychology*, vol. 107, no. 1, pp. 135–153, 2016.

[31] R Core Team, "R: a language and environment for statistical computing," 2013, https://www.R-project.org/.

[32] W. Revelle, "psych: procedures for psychological, psychometric, and personality research," 2016, May 2017, https://cran.r-project.org/web/packages/psych/index.html.

[33] L. Van Der Krieke, B. F. Jeronimus, F. J. Blaauw et al., "HowNutsAreTheDutch (HoeGekIsNL): a crowdsourcing study of mental symptoms and strengths," *International Journal of Methods in Psychiatric Research*, vol. 25, no. 2, pp. 123–144, 2016.

[34] S. Epskamp, M. K. Deserno, and L. F. Bringmann, "mlVAR: multi-level vector autoregression," R package version 0.4, 2017, https://CRAN.R-project.org/package=mlVAR.

[35] S. Epskamp, L. J. Waldorp, R. Mõttus, and D. Borsboom, "Discovering psychological dynamics: the Gaussian graphical model in cross-sectional and time-series data," 2016, https://arxiv.org/abs/1609.04156v2.

[36] G. Csárdi and T. Nepusz, "The igraph software package for complex network research," *InterJournal, Complex Systems*, vol. 1695, 2006.

[37] S. Epskamp, A. O. J. Cramer, L. J. Waldorp, V. D. Schmittmann, and D. Borsboom, "Qgraph: network visualizations of relationships in psychometric data," *Journal of Statistical Software*, vol. 48, no. 4, 2012.

[38] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, 2015.

[39] M. J. Rovine and T. A. Walls, "Multilevel autoregressive modeling of interindividual differences in the stability of a process," in *Models for Intensive Longitudinal Data*, Oxford University Press, 2006.

[40] E. L. Hamaker and R. P. P. P. Grasman, "To center or not to center? Investigating inertia with a multilevel autoregressive model," *Frontiers in Psychology*, vol. 5, 2015.

[41] S. W. Raudenbush and A. S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, SAGE, 2002.

[42] K. Bulteel, F. Tuerlinckx, A. Brose, and E. Ceulemans, "Using raw VAR regression coefficients to build networks can be misleading," *Multivariate Behavioral Research*, vol. 51, no. 2-3, pp. 330–344, 2016.

[43] N. K. Schuurman, E. Ferrer, M. de Boer-Sonnenschein, and E. L. Hamaker, "How to compare cross-lagged associations in a multilevel autoregressive model," *Psychological Methods*, vol. 21, no. 2, pp. 206–221, 2016.

[44] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[45] E. Snippe, W. Viechtbauer, N. Geschwind, A. Klippel, P. de Jonge, and M. Wichers, "The impact of treatments for depression on the dynamic network structure of mental states: two randomized controlled trials," *Scientific Reports*, vol. 7, no. 1, article 46523, 2017.

[46] L. K. Muthén and B. O. Muthén, *Mplus User's Guide*, Muthén & Muthén, Los Angeles, CA, USA, 1998.

[47] M. A. Cohn, B. L. Fredrickson, S. L. Brown, J. A. Mikels, and A. M. Conway, "Happiness unpacked: positive emotions increase life satisfaction by building resilience," *Emotion*, vol. 9, no. 3, pp. 361–368, 2009.

[48] B. L. Fredrickson, "What good are positive emotions?," *Review of General Psychology*, vol. 2, no. 3, pp. 300–319, 1998.

[49] F. Bryant, "Savoring Beliefs Inventory (SBI): a scale for measuring beliefs about savouring," *Journal of Mental Health*, vol. 12, no. 2, pp. 175–196, 2003.

[50] T. B. Kashdan and J. Rottenberg, "Psychological flexibility as a fundamental aspect of health," *Clinical Psychology Review*, vol. 30, no. 7, pp. 865–878, 2010.

[51] P. Kuppens, L. B. Sheeber, M. B. H. Yap, S. Whittle, J. G. Simmons, and N. B. Allen, "Emotional inertia prospectively predicts the onset of depressive disorder in adolescence," *Emotion*, vol. 12, no. 2, pp. 283–289, 2012.

[52] A. Klippel, W. Viechtbauer, U. Reininghaus et al., "The cascade of stress: a network approach to explore differential dynamics in populations varying in risk for psychosis," *Schizophrenia Bulletin*, vol. 44, no. 2, pp. 328–337, 2018.

[53] B. H. Strand, O. S. Dalgard, K. Tambs, and M. Rognerud, "Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36)," *Nordic Journal of Psychiatry*, vol. 57, no. 2, pp. 113–118, 2003.

[54] L. R. Demenescu, A. Stan, R. Kortekaas, N. J. A. van der Wee, D. J. Veltman, and A. Aleman, "On the connection between level of education and the neural circuitry of emotion perception," *Frontiers in Human Neuroscience*, vol. 8, 2014.

[55] E. B. McClure, "A meta-analytic review of sex differences in facial expression processing and their development in infants,

children, and adolescents," *Psychological Bulletin*, vol. 126, no. 3, pp. 424–453, 2000.

[56] S. Nolen-Hoeksema and A. Aldao, "Gender and age differences in emotion regulation strategies and their relationship to depressive symptoms," *Personality and Individual Differences*, vol. 51, no. 6, pp. 704–708, 2011.

[57] R. P. Bagozzi, N. Wong, and Y. Yi, "The role of culture and gender in the relationship between positive and negative affect," *Cognition and Emotion*, vol. 13, no. 6, pp. 641–672, 1999.

[58] M. Houben, W. Van Den Noortgate, and P. Kuppens, "The relation between short-term emotion dynamics and psychological well-being: a meta-analysis," *Psychological Bulletin*, vol. 141, no. 4, pp. 901–930, 2015.

[59] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[60] P. Koval, M. L. Pe, K. Meers, and P. Kuppens, "Affect dynamics in relation to depressive symptoms: variable, unstable or inert?," *Emotion*, vol. 13, no. 6, pp. 1132–1141, 2013.

Hindawi
Complexity
Volume 2018, Article ID 6438702, 14 pages
https://doi.org/10.1155/2018/6438702

WILEY | Hindawi

Research Article

Cohort and Rhyme Priming Emerge from the Multiplex Network Structure of the Mental Lexicon

Massimo Stella

Institute for Complex Systems Simulation, University of Southampton, Southampton, UK

Correspondence should be addressed to Massimo Stella; massimo.stella@inbox.com

Received 16 May 2018; Accepted 9 August 2018; Published 17 September 2018

Academic Editor: Cynthia Siew

Complex networks recently opened new ways for investigating how language use is influenced by the mental representation of word similarities. This work adopts the framework of multiplex lexical networks for investigating lexical retrieval from memory. The focus is on priming, i.e., exposure to a given stimulus facilitating or inhibiting retrieval of a given lexical item. Supported by recent findings of network distance influencing lexical retrieval, the multiplex network approach tests how the layout of hundreds of thousands of word-word similarities in the mental lexicon can lead to priming effects on multiple combined semantic and phonological levels. Results provide quantitative evidence that phonological priming effects are encoded directly in the multiplex structure of the mental representation of words sharing phonemes either in their onsets (cohort priming) or at their ends (rhyme priming). By comparison with randomised null models, both cohort and rhyming effects are found to be emerging properties of the mental lexicon arising from its multiplexity. These priming effects are absent on individual layers but become prominent on the combined multiplex structure. The emergence of priming effects is displayed both when only semantic layers are considered, an approximated representation of the so-called semantic memory, and when semantics is enriched with phonological similarities, an approximated representation of the lexical-auditory nature of the mental lexicon. Multiplex lexical networks can account for connections between semantic and phonological information in the mental lexicon and hence represent a promising modelling route for shedding light on the interplay between multiple aspects of language and human cognition in synergy with experimental psycholinguistic data.

1. Introduction

Cognitive network science is quickly rising as an interdisciplinary field exploring psychology with the quantitative tools derived from complex networks [1–4]. Through the lens of network theory, many recent studies investigated the cognitive representation of language, a system commonly called mental lexicon [5] and deeply influencing processes such as language learning [6–10], memory [11–14], creativity [3, 15], and language decline in cognitive impairments [16–18]. It has to be underlined that these studies are only a small part of a much wider literature on the mental lexicon from psycholinguistics [5, 19–21].

Network science provided language scientists with quantitative ways of representing and analysing the structure of lexical items within the mental lexicon [1, 4, 12, 22]. For instance, concepts such as percolation techniques were used for detecting patterns of word confusability in phonology [12, 22], strategies of language learning in healthy and clinical populations of children [6, 23], differences in the levels of creativity of individual healthy subjects [3, 11], or differences in the production of words in people with aphasia [17, 18] or Alzheimer's disease [24]. However, the above studies considered only one aspect of language for establishing similarities among words, e.g., building single-layer networks including only phonological similarities among words [12]. While this focus was valuable for investigating on large scales how thousands of similarities among words influenced processes such as word identification or memorisation tasks [12, 13, 25], the way humans store and memorise words is inherently multirelational [1, 23, 26]. Multiple types of semantic and phonological similarities among words are present simultaneously, and they can either compete or assist specific language processes in different ways [1, 5, 27, 28]. For

instance, a recent empirical investigation indicated that toddlers simultaneously exploit both phonological and semantic features of words in early language learning [23, 29].

Phonological and semantic relationships can also affect lexical retrieval in different ways. Lexical retrieval is a set of cognitive processes and executive functions related to the identification of a specific cognitive unit (e.g., a word) from semantic memory [30] subsequently to a given visual or auditory input (e.g., hearing or reading a given word) [15, 31–33]. Conceptual similarities can cause the so-called *priming*-phenomenon, where one lexical item (a prime) facilitates or inhibits the retrieval of another word (a target) [5, 32, 34, 35]. Priming can happen with different modalities depending on how prime and target are processed (e.g., visual-visual, auditory-auditory, or crossmodal) and can involve perceptual, semantic, or conceptual types of similarities between prime and target [35]. Facilitative semantic priming happens when a target word (e.g., "hawk") is processed faster and more accurately when preceded by a semantically related stimulus (e.g., "dove") than when preceded by an unrelated word (e.g., "prosthetics") [34]. Empirical work has shown that facilitative semantic priming decayed more quickly over time when words were processed individually compared to when words were processed in sentences [35, 36]. This empirical evidence has been linked with the richer structure of semantic associations among words in a sentence [35, 36], indicating a positive correlation between word-word associations and facilitatory semantic priming. On the other hand, semantic inhibition or interference happens mainly through visual and perceptual modalities [35]. For instance, ignoring a picture representing a "dog" can produce subsequent slowing when responding to the word "cat".

Semantic priming typically only considers primes and targets belonging to the same semantic category (e.g., "hawk" and "dove" are both types of birds). However, words can be semantically related in other ways, which were often captured through free associations (e.g., "bed" and "pillow" are often provided as free associations when talking about bedroom furniture). Indeed, associative priming has been shown to crucially depend on the time between the beginning of the prime and the onset of the target [35, 37], a time window also called *stimulus onset asynchrony* (SOA). A longer SOA between prime-target pairs corresponded to stronger facilitative priming effects, whereas nonassociated prime-target pairs corresponded to inhibitory priming effects independent to the SOA. Rather than exploiting taxonomical, semantic, or cooccurrence similarities, perceptual priming depends on the form of the stimulus. A similar priming effect occurs with phonological similarities [28, 38]. Hearing primes can lead to easier lexical processing of phonologically similar target words [28, 35].

Inhibitory priming relies on mechanisms restricting access to specific concepts, and the investigation of such inhibitory dynamics still represents an open challenge in the relevant literature [5, 28, 35]. Facilitative priming is well explained by network models of semantic memory [5, 15, 28, 35, 39, 40] using spreading activation mechanisms. Although its mechanisms remain an open challenge in neuropsychology [5, 11], past attempts have successfully modelled semantic memory as a complex network in order to obtain limited but meaningful insights of facilitative priming effects and lexical retrieval latencies in word identification tasks [1, 2, 15, 39, 40]. Collins and Loftus represented semantic memory as a conceptual network with links placed between concepts that shared features. When a given stimulus was activated (e.g., reading the word "animal"), then many words in the semantic levels of the mental lexicon received portions of activation, proportionally to their semantic relatedness to the stimulus. The activation spread across semantic similarities and it ensued until it converged on a single target, more or less related to the stimulus, which was then retrieved. Hence, lexical retrieval of an item was relative to a network node receiving a convergence of activation from across its connections. Importantly, the spread of activation could cover far distances of time but decreased in intensity. According to this model, the retrieval of target words was facilitated by having primes close or adjacent to the prime words. Furthermore, the model could interpret empirical evidence of longer SOAs leading to stronger facilitative priming [37] in terms of activation accumulating over a given lexical item, leading to faster and more accurate concept retrieval.

In Collins and Quillian's experiments [40], subjects were asked to read and verify statements relating to two concepts, e.g., *a canary is a bird*. The time it took for participants to verify a statement correlated positively with the distance between concepts (e.g., canary and bird) in the conceptual network representation of semantic memory [39, 40], i.e., the smallest number of semantic similarities connecting concepts. This represented preliminary evidence that network distance in semantic networks correlates with lexical retrieval patterns, although it was limited only to a rigid network structure encompassing only semantic features of words.

More recent approaches have modelled a semantic network as a web of free associations among concepts [3, 11, 15], i.e., relationships based on memory rather than on any strict definition of feature sharing. The importance of network distance for quantifying patterns of lexical retrieval was recently underlined in the recent work by Kenett et al. [15]. The authors showed that success in free- and cued-recall experiments decreased dramatically with increasing distance between concepts in a network of free associations. Furthermore, network distance predicted success in recall experiments considerably better than mainstream psycholinguistic techniques such as latent semantic analysis [34]. Network distance has also been shown to influence lexical retrieval when considering a phonological network. For example, recent investigations showed how words at shorter mean network distance were more promptly recognised in a lexical decision task [14, 25]. These results strongly indicate a cognitive advantage in processing concepts at shorter network distances. In a spreading activation model of lexical retrieval, network distance might capture how spreading activation decays over the mental lexicon structure, further promoting the usage of network models and network distances for the investigation of lexical retrieval.

Additional empirical evidence has shown that phonological similarities can reduce naming latencies in picture

naming tasks, an effect known as phonological facilitation [27]. This evidence led to the inclusion of phonological aspects of the mental lexicon for obtaining more refined models of lexical retrieval from the auditory input. In case of hearing a word rather than reading it, more recent work has proposed a spreading activation mechanism including phonological similarities among words [12, 41–44]. Within a bottom-up process, activation first spreads among phonological neighbours of the stimulus and then moves up across semantic memory, ultimately leading to word identification and retrieval.

In agreement with the above approaches, the present study adopts the assumption that the mental lexicon encapsulates not only linguistic features of individual words (e.g., their meaning, their orthography, their phonology, etc.) but also their similarities. However, the present investigation builds on the previous network approaches to lexical retrieval [14, 15, 25] by considering within the same network representation both semantic and phonological similarities among words through the framework of multiplex lexical networks [8–10, 16, 45]. In a multiplex lexical network, nodes represent words and links connect words differently according to specific network layers of similarities [8, 9, 45]. For instance, Stella et al. [8, 10] used a multiplex lexical network with layers representing free associations, shared semantic features, cooccurrences, and phonological similarities, which successfully predicted early word acquisition in toddlers. The first large-scale application of multiplex lexical networks was from Stella et al. [9], where the mental lexicon of an adult was approximated as a multilayer network with four layers of word similarities: free associations, synonyms, generalisations, and phonological similarities. Through a data-driven approach, intersecting many large-scale datasets about word frequency, age of acquisition, concreteness, and reaction times in lexical identification tasks, the authors identified a multiplex lexical core, a set of words tightly interconnected with each other, appearing suddenly during normative development around age 8 yrs. This core made the whole multiplex lexical network extremely resilient to cognitive impairments modelled as progressive random word removal. Multiplex lexical networks were adopted also in a clinical population of people with aphasia, revealing the importance of the multiplex structure for predicting correct picture naming [16].

This paper adopts multiplex lexical networks for studying two specific patterns of phonological priming in lexical retrieval: cohort priming and rhyme priming. The term *cohort priming* comes from cohort theory, a theory of lexical retrieval by Marslen-Wilson and colleagues [31]. When hearing speech, the first phoneme heard "activates" every word in the lexicon with that phoneme in an access stage, resulting in a "cohort of words". For instance, hearing *belief* initially activates all words starting with the phoneme /*b*/, resulting in a very large cohort of possible words. As the next phoneme is heard, the cohort is further restricted, in this case, to words starting with /bI/ and so on, phoneme by phoneme. As more phonemes are added, fewer and fewer words are found as candidates until a recognition point is reached such that only one word is activated [31, 33]. This recognition point is known also as *isolation point* or *uniqueness point* [31]. Cohort theory assumes a quite strict definition of cohorts and it does not consider lexical effects due to the structure of word-word similarities in the heard input (e.g., phrasal context) or in the mental lexicon [33]. However, empirical studies have confirmed that the initial portion of a word activates similar sounding words that compete for recognition and, more importantly, are quicker to identify when primed by words in the same cohort [31, 33, 46]. This facilitatory *cohort priming* effect was detected in case either primes were English words or nonwords sharing the first three phonemes with the target [46], supporting the assumption of activation of lexical items based on their initial phonetic structure. Notice that the simultaneous activation of lexical items corresponds not only to facilitatory priming effects but also to lexical competition in distinguishing words from the same cohorts [47]. In word identification tasks without priming, targets in larger cohorts were found to be recognised less accurately than targets in smaller cohorts [47]. However, this competition effect disappeared when words were presented in a phrasal context [28], indicating that the semantic and syntactic features of words extracted by sentences can interact with cohort structure and influence lexical retrieval of words in cohorts. The above experimental findings motivate further investigation of cohort priming effects also in relation to the semantic and syntactic levels of the mental lexicon.

Rhyme priming is analogous to cohort priming, in that sharing phonemes at the end of words can give rise to facilitatory priming effects [46]. According to the relevant literature of priming effects, primes rhyming with a target lead to shorter and more accurate lexical retrieval compared to nonrhyming primes [46]. A similar rhyme facilitation of lexical decisions to real-world targets was found also in nonfluent people with aphasia [48]. Rhyming priming also has beneficial effects for the memorisation of words [49], especially in young children [38]. Empirical studies have shown that this type of priming is weaker than cohort priming but still present during lexical retrieval [49]. The current investigation of cohort and rhyme priming differs substantially from previous analyses of cohort and rhyme priming. Here, by assuming a network representation of the semantic and phonological subcomponents of the mental lexicon, the main aim is to detect cohort and rhyme priming effects in thousands of words by harnessing directly the structure of dozens of thousands of word-word similarities of different types rather than directly testing only a limited number of words, as in previous lab experiments [31, 33, 38, 49]. This multiplex network approach has three main strengths: (i) it can quantify which semantic or phonological layers are predominantly involved in potential priming effects; (ii) it can account for any potential interplay and nonlinear effects over priming arising from combining semantics and phonology, an interplay often neglected in previous network studies; (iii) it can be performed at large scales, testing a sample of words up to two orders of magnitude larger than in previous lab experiments [47].

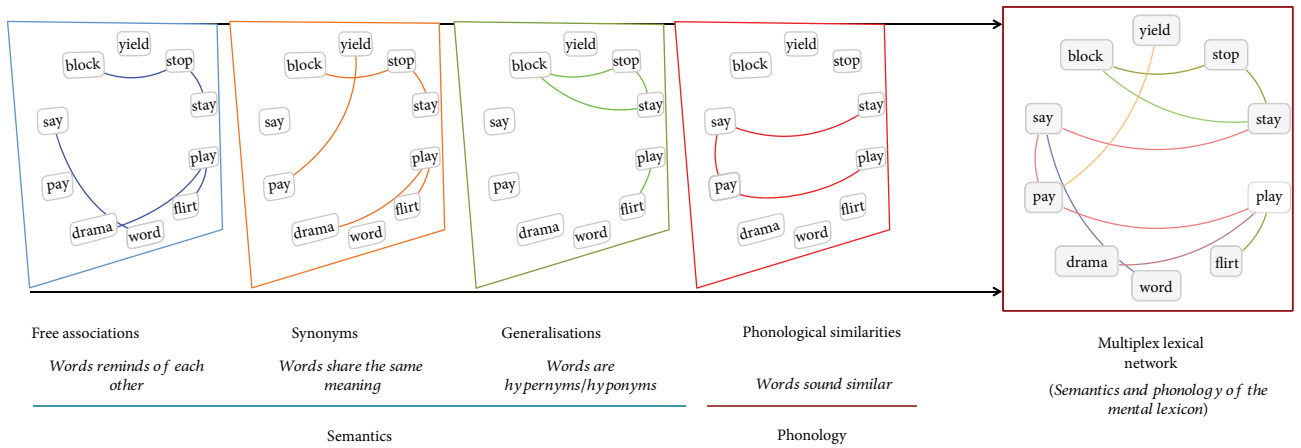


Figure 1: Network visualisation of a portion of the adopted multiplex lexical network. The whole multiplex representation contains 8546 words. Semantic layers are clustered together (free associations, synonyms, and generalisation) and represent multiple aspects of semantic memory. Phonological information is represented as a network of phonological similarities, where words differing by one phoneme were linked together. The resulting multiplex lexical network is an edge-coloured graph where links of different types coexist (see right panel).

2. Methods and Model

This section provides information on (i) the construction of the multiplex lexical network, (ii) the linguistic datasets used, (iii) the network metrics adopted and their psycholinguistic interpretation, and (iv) the null models used as a reference.

2.1. Construction of the Multiplex Lexical Network. The mental lexicon of an adult English speaker was represented as a multiplex lexical network including 8546 words connected over four network layers, analogous to previous approaches [9, 16]. The layers have been selected according to the spreading activation model for auditory input [12, 18, 21, 41, 42], in which language processing happens first over a subcomponent containing phonological information about words and subsequently over semantic memory. Hence, the multiplex lexical network is chosen in order to combine phonological and semantic aspects of language. More in detail, information about phonology is mediated by a layer of phonological similarities [4, 22], where words are connected if they differ in the addition/substitution/deletion of one phoneme, e.g., "cat" would be connected to "cab" because of the above operational definition of sound similarity. Notice that other patterns of sound similarity are not directly captured by this metric (e.g., "cat" and "cob", which are 2 phoneme substitutions apart). Information about semantic memory is encapsulated within three different levels:

(i) overlap in meaning was encapsulated in a layer of synonyms, where words were connected if they can have the same meaning, e.g., "meaningful" and "insightful" can have the same meaning

(ii) the linguistic hierarchy of concepts was encapsulated in a layer of generalisations, where words were connected if they belonged to either a more specific or a more general semantic category, e.g., "dove" is a type of "bird"

(iii) most of the remaining semantic similarities among words were encapsulated within a layer of empirical free associations, where words were connected if they were associated by participants during a free association tasks, e.g., "bed" reminds participants of "sleep"

It is important to underline that free associations, generalisations, synonyms, and phonological similarities were all found to deeply affect lexical retrieval in several independent studies [1, 2, 12, 44, 50], hence the importance of including them in the current investigation. The free association network was built as a subgraph of the Edinburgh Associative Thesaurus [50]. The synonym, the generalisation, and the phonological networks were built according to a dataset managed by Wolfram Research and based on WordNet 3.0 [51]. All layers were treated for simplicity as undirected, and no cost associated with between-layer transitions was considered, analogous to previous studies in the relevant literature [8, 15, 16]. Word features such as frequency were obtained from the large-scale repository Opensubtitles [52], which computes word frequencies from subtitles in TV series and movies.

As reported in Figure 1, the resulting multiplex network represents an edge-coloured graph [53, 54]. The same set of nodes is replicated on each layer but different types/colours of links among nodes can be present, with each colour corresponding to a specific layer. On this structure, transitions between layers are allowed by transitioning between replicas of nodes. The multiplex structure alters dramatically the layout of similarities among words. Words disconnected on a layer might be highly connected and central on the whole multiplex structure, like for instance "say" in the layer of generalisations and in the whole multiplex lexical network (see Figure 1).

The imbalance in modelling the multiplex lexical network with three semantic layers but only one phonological layer is due to (i) the relative importance in distinguishing

different semantic aspects of the lexicon (e.g., synonyms are different from taxonomical relationships) and (ii) to the relative difficulty of considering measures of sound similarities that provide more information than the definition of phonological similarity adopted in this work (cf. [4]). However, it should be noted that the free association layer overlaps more than random expectation with the layer of phonological similarities [8], indicating that the association layer is not purely semantic but it contains also some sort of phonological information in it. This reduces the imbalance between semantics and phonology in the chosen representation. Nonetheless, previous similarity results [9] indicated that the layer of free association still contains patterns of word-word similarities that were more similar to those encoded in the synonym and generalisation layers rather than to the phonological layer. For the present analysis, the free association layer was considered as a semantic layer, compatible with what previous studies assumed [2, 3, 15, 55].

2.2. Testing Cohort Theory. According to the cohort model, lexical retrieval happens when the isolation point (see Introduction) is reached, corresponding to a peak time inactivation [31, 33]. Phonemes heard prior to the peak time determine the onset of the word and, consequently, the number of words in that word's cohort. While the peak time may change for each word based on its context, empirical evidence indicates that the average peak time of a word is around 200 ms from when the word gets pronounced [31] and corresponds to having information about the first 3 or 4 phonemes of the word [28, 33]. Note that the above numbers represent average estimations, since the number of phonemes occurring in the 200 ms window can vary depending on the phoneme types (e.g., stops vs. fricatives vs. nasals). Since in the current dataset considering onsets made of 4 phonemes led to quite small cohorts, the focus shifted on onsets made of 3 phonemes, as tested also in previous studies [46]. For every onset available in the current dataset, a cohort of words was built. In order to reduce the extent of systematic errors due to small sample sizes, only cohorts with more than 10 words were considered. This led to the selection of 2526 words from the multiplex lexical network. Selected words were subdivided into 99 cohorts of average size 30 ± 10 words.

2.3. Testing Rhyme Priming. Rigid definitions like considering only the overlap in phonemes in the last positions of words cannot capture the wide variety of rhyming patterns in English [49]. Rhymes depend not only on phoneme structure but also on additional features, like stress. In order to overcome this issue, the online rhyming dictionary RhymeZone was used for selecting groups of rhyming words [56]. RhymeZone is partially based on WordNet [51] but it is also enriched with additional data from quotes and lyrics. The complete corpus of RhymeZone includes semantic and phonological information over almost 19 million words from 1061 dictionaries; hence, it represents a large-scale and cross-checked source of current rhymes in the English language. The current analysis focused on true rhymes, i.e., words with identical sounds after a stressed vowel. Homophones, different words having exactly the same phonemes, were not considered as rhyming words. According to this choice, 2247 rhyming words were selected from the multiplex lexical network. Selected words were subdivided into 51 rhyme classes (e.g., all words rhyming with "authorisation"), of average size 40 ± 10 words. In order to reduce the extent of systematic errors due to small sample sizes, only classes with more than 10 words were considered.

2.4. Network Metrics. As indicated in many recent investigations about lexical retrieval in semantic and phonological subcomponents of the mental lexicon, network distance is a reliable proxy of word relatedness as it is predictive of lexical retrieval [3, 11, 15, 57]. Network distance d_{ij} between nodes i and j in a given network N is defined as the shortest number of links connecting i and j [58]. In cases where there is no path connecting i and j, then nodes i and j are said to be disconnected and d_{ij} is assumed to be equal to ∞. As reported in Figure 1, in the multiplex lexical network, paths can be made of links of different layers/colours. Therefore, there can be additional, nontrivial "multiplex" paths emerging from the multiplex structure, so that the network distance between two words on any individual layer can be dramatically different from the network distance between the same words on the whole multiplex network. For instance, *bed* and *sleep* might be disconnected on the phonological layer but connected on the free association layer. This richer behaviour of network distance on the multiplex network represents the interplay between phonological and semantic aspects of the mental lexicon. Notice that the whole multiplex lexical network is fully connected in the sense of De Domenico et al. [54], i.e., there is always a multiplex path connecting any two words when transitions across layers are allowed. However, individual layers are not fully connected, so that some words might be disconnected and hence correspond to a divergent distance $d_{ij} = \infty$. In order to overcome the issue of having infinite distances, the closeness c_{ij} of nodes i and j [58] is used, namely, the inverse of network distance:

$$c_{ij} = \frac{1}{d_{ij}}, \tag{1}$$

where $c_{ij} = 0$ when i and j are disconnected. Considering the inverses of network distance gets rid of divergences, so that average finite estimators of distance can be computed. Provided that in the analysis individual network layers might be disconnected, a valid proxy for the central moment of the distribution of closeness is represented by the mean [58]:

$$c^* = N \sum_{i \neq j} \frac{1}{d_{ij}}, \tag{2}$$

ranging from 0 (all nodes are disconnected) to 1 (all nodes are adjacent with each other). c^* represents the harmonic mean of the distances of all node pairs in a given network, a measure also called efficiency [58]. Notice that c^* is analogous but not equivalent to closeness centrality C_i, which is the arithmetic mean of distances of node pairs (for a

comparison see [58]). In disconnected networks, the harmonic mean is a better estimator of closeness compared to the arithmetic mean; hence, in the following, c^* is adopted for estimating how close words are on the multiplex lexical network. We assume that primes and targets that are closer on a network topology are processed faster and more accurately than words at greater network distance, as supported by recent empirical studies [15, 25, 57]. Closeness is computed among words in specific subsets: (i) words in the same cohort and (ii) words having the same rhyme (i.e., composing a rhyming class).

2.5. Null Models. Quantifying the average closeness of words in cohorts and in rhyme classes requires a suitable null model for comparison and statistical testing. Since phonological information is important for defining both cohorts and rhymes, considering randomised lists of words satisfying constraints at the phonological level is an intuitive choice. As a viable approach, randomised cohorts/rhyme classes are built by sampling at random real words sharing at least m phonemes in any position. Both consecutive and nonconsecutive shared phonemes had to be considered, since limiting the null model to consider only overlapping consecutive phonemes outside of the onset/end resulted in sample size issues, e.g., too few words for statistical comparisons with cohorts and rhyme classes. Randomised cohorts/classes have the same size of the original ones. For cohorts, m is equal to, because in the operative definition of cohorts onsets are defined as having the same first three phonemes as a consequence of the average peak time. For rhymes, m can range between 2 and 4; the appropriate value is computed by calculating the number of phonemes that all words in a rhyme class have at their ends. The same m phonemes defining a cohort/class are used for building its randomised counterpart. For instance, consider the cohort "belief", "belong", "beloved", ... defined by phonemes /b/, /I/, /l/. A randomised cohort will include words sharing these phonemes but in positions different from the onset, e.g., "automobile", "abolish", "assembly", ... Preserving phoneme identity is important because different phonemes might lead to differences in phonological awareness and influence lexical processing [20].

The phonological constraint on the randomised lists guarantees that the same phonemes are present in both original cohorts/rhyme classes but in positions different from the onset/end of the word. Therefore, the considered null models allow us to test how phoneme sequences at the beginning and at the end of individual words influence lexical processing in relation to the multiplex structure. Hence, the proposed methodology investigates to what extent the multiplex lexical network is nonrandomly structured to cluster onset-sharing and rhyme-sharing words. To this aim, differences among individual phoneme sequences are averaged across cohorts/rhyme classes and a statistical test is performed between the average closeness of cohorts/rhyme classes and random expectation from the above null models. Nonparametric statistical testing, specifically a sign test, is adopted in order to obtain results robust to violations of normality due to the low sample size of cohorts or rhyme classes. On the layers of free associations, synonyms, and generalisations, the distribution of average closeness for words in cohorts and rhyme classes was found to violate normality (Kolmogorov-Smirnov test, $D > 0.08$,$p > 0.09$) at a 0.05 significance level.

Comparison with the null models also enables one to test whether potential differences in closeness between cohorts/rhymes and the randomised lists can be explained either by individual aspects of language or by the interplay between them, e.g., phonology and semantics or different aspects of semantics. This is achieved by computing network distance on individual layers and on the whole multiplex network representation separately. These results are then compared against another set of null models for the network layers where links are randomised. In each randomised layer, words have the same number of connections as in the respective empirical layer but connections are rewired uniformly at random. Hence, random rewiring preserves the degree distribution of words on a layer. Since the same word can have different degrees on different layers [9], then different rewired null models have to be adopted for the different layers of the multiplex network. These null models are also called configuration models in the network literature [59], and they preserve the number of total word-word similarities of individual words (i.e., nodes degrees) and also the heterogeneity in the number of similarities individual words can have (i.e., degree heterogeneity). Randomly rewiring every individual layer is expected to disrupt both intralayer correlations between nodes and interlayer correlations between links. Therefore, configuration models allow quantifying to what extent differences in closeness between cohorts/rhyme classes and random lists of words are due to either global patterns of network structure (which are disrupted by random rewiring) or just by heterogeneity in link allocation (which is fixed even under random rewiring).

3. Results

Results are presented in two stages. First, the suitability of the adopted representation from a language perspective (considering word frequency) and from a network perspective is reported. Cohort and rhyme priming effects are then analysed by using network distance and by considering specific reference null models as a comparison.

3.1. The Relevance of the Multiplex Lexical Representation. The selected multiplex network representation of the mental lexicon is composed of layers including semantic and phonological aspects of the mental lexicon of relevance in the literature about lexical retrieval (see also Methods). However, this structure needs further validation since it must: (i) correspond to commonly used words, and also (ii) correspond to a structure that cannot be further aggregated, i.e., network layers should display different patterns of word similarities in order to further motivate the choice of considering them as separate multiplex layers.

Figure 2 reports the frequencies of words in the multiplex lexical network and in reference datasets from Opensubtitles [52]. The probability of finding words with a frequency

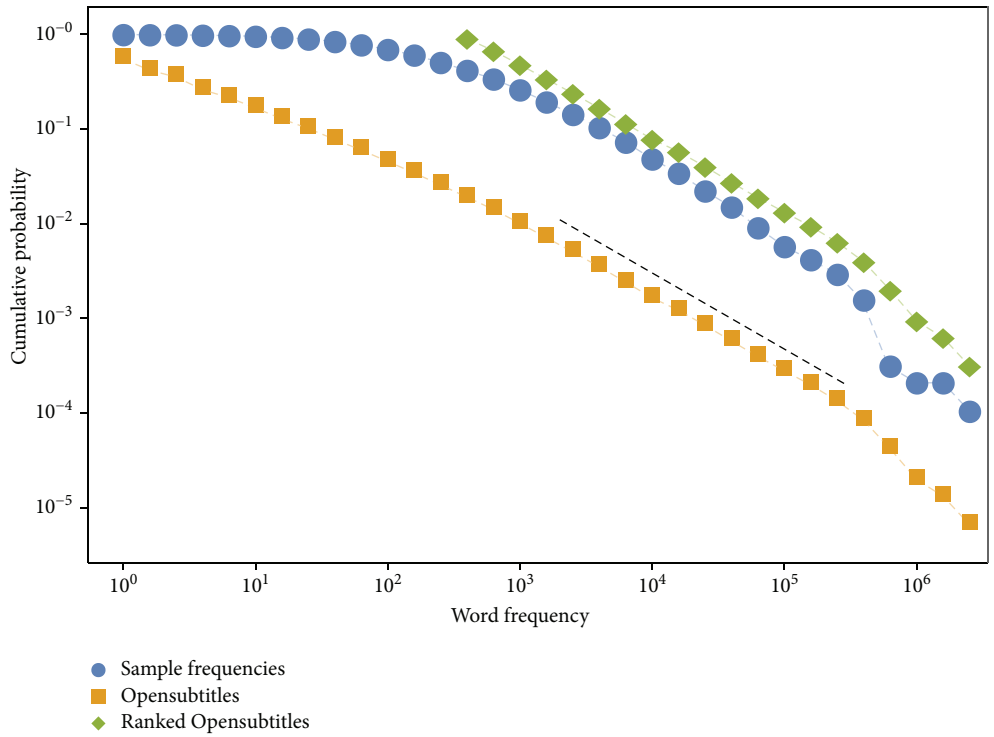


Figure 2: Cumulative probability distribution $P(X \geq f)$ of finding a word with frequency at least f in the multiplex lexical network with $N = 8531$ words (blue dots), in the Opensubtitles dataset with $5{\cdot}10^5$ unique words (orange squares), and in the subset of $N = 8531$ most frequent words from Opensubtitles (green diamonds). A power-law with exponent −1.8 is reported for visual comparison (dashed line).

higher than 10 is one order of magnitude larger in the multiplex network than in the whole Opensubtitles dataset. The multiplex lexical network is richer in terms of commonly used words compared to the language used in movies, which can also contain more specific and less frequent words (e.g., specific jargon, geographical names, etc.). Furthermore, the words in the multiplex lexical network are almost as frequent as the most frequent words in Opensubtitles (see for reference the probabilities of finding words with a frequency higher than 10^3 in Figure 2). Based on these results, the conclusion is that, in terms of word frequency, the multiplex lexical network includes commonly used words and is the representative of the most common semantic and phonological features of spoken English.

The choice of keeping free associations, synonyms, generalisations, and phonological similarities as separate is supported by a structural reducibility analysis, an entropy-based technique for establishing the information about network paths that is lost when layers are aggregated in a given multiplex network (see De Domenico et al. [60] for the technical details). Analogous to previous investigations with multiplex lexical networks based on other datasets [8, 45], the multiplex lexical network used in the current study (cf. [9]) is irreducible. In other words, a significant number of patterns of word-word similarities could be lost in case any two or more layers of the multiplex lexical network were projected onto one layer only. The free association layer is also found to be distinct compared to generalisation, synonyms, and phonological similarities, so that it should not combine with any of these three layers. This finding confirms that the considered layers are representative of different aspects of the mental lexicon, which should be kept as distinct.

All in all, the frequency analysis indicates that the investigated multiplex lexical network is almost as rich in commonly used English words and poorer in terms of more infrequent lexical items when compared to the larger sample of words from Opensubtitle, which includes with $5{\cdot}10^5$ lexical items and is the representative of currently spoken English. The irreducibility analysis is another important element as it motivates the consideration of the chosen aspects of semantics and phonology through separate layers in the multiplex network. Hence, both the frequency and the structural reducibility analyses confirm the suitability of the multiplex lexical representation for investigating patterns of the mental lexicon for the English language.

3.2. The Multiplex Lexical Network Identifies Cohort Priming. As reported in the introduction, facilitative semantic, associative, and phonological priming effects are well explained by activation spreading models over shorter network paths of word-word similarities in the mental lexicon [35, 37, 39, 41]. There is additional evidence that also inhibitory semantic priming depends on the proximity of concepts on semantic networks [61]. Furthermore, as confirmed by recent studies [15, 25, 57], closeness is a reliable estimator of the efficiency of lexical processing; closer words on semantic and phonological networks tend to be retrieved faster and more accurately than words farther apart.

Figure 3(a) compares the median closeness of cohorts (orange bars) and of random lists (blue bars) on individual

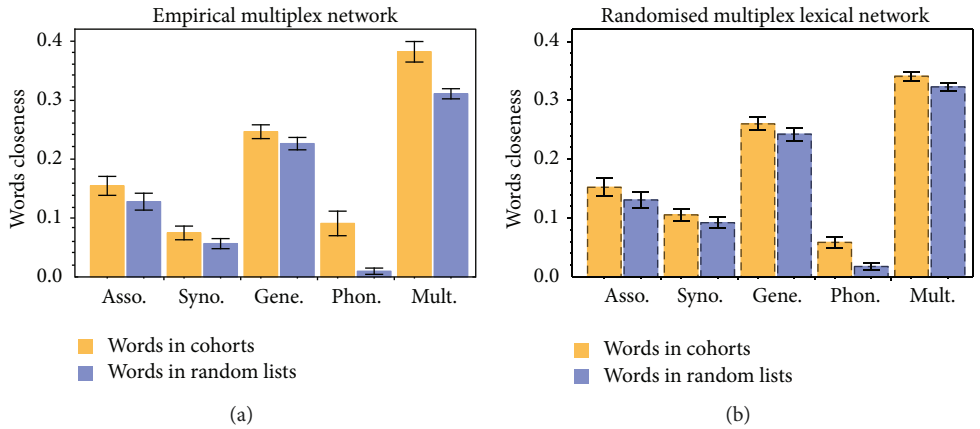


FIGURE 3: (a) Mean closeness distance of words either in cohorts or in randomised lists in the layers, respectively, made of free associations (Asso.), synonyms (Syno.), generalisations (Gene.), and phonological similarities (Phon.). (b) The same as in (a) but for randomised layers of word-word similarities.

layers and over the whole multiplex structure. Error bars indicate error margins over the median. At the significance level $\alpha = 0.05$, the differences in closeness between words in cohort and words in null models are not statistically significant on the free association layer (sign test, $n_+ = 52$, $p = 0.65$) and on the synonyms layer (sign test, $n_+ = 56$, $p = 0.23$). Statistically significant differences are observed on the generalisations layer (sign test, $n_+ = 72$, $p < 10^{-5}$) and on the phonological layer (sign test, $n_+ = 63$, $p = 0.008$). A statistically significant difference is found also on the whole multiplex structure (sign test, $n_+ = 73$, $p < 10^{-5}$).

Words in cohorts are found to be on average closer than random expectation on specific layers, indicating the presence of a cognitive influence when processing them together and hence a priming effect. The gap observed in the phonological layer can be attributed to a tendency for words in the same cohorts to persist in the same connected component. In fact, the lower inverse network distance/closeness of the null model relates with the fragmentation of the phonological network (cf. [22]), so that words in the same cohort can have zero closeness, and this ultimately lowers the average closeness score. Therefore, despite both phonological links and cohorts being based on measures of phonological similarities, the observed gap between empirical and random average closeness of words is an indication of the clustering of cohorts over the same connected components in the multiplex lexical network. Interestingly, also cohorts in the generalisation layer are closer than random expectation. Provided that cohorts are based on word forms, this clustering over a semantic layer might be the consequence of a form-meaning correlation, a phenomenon called form-meaning nonarbitrariness and empirically traced in English and many other languages [62]. The magnitude of the gaps in closeness found over the multiplex network and over the generalisations and the phonological layer do not correlate with the cohort size (Kendall Tau $|\tau| < 0.07$, p values > 0.4). This analysis is directly based on the layout of hundreds of thousands of word similarities in a multiplex lexical network representative of commonly spoken language.

Notice that the difference in closeness between cohorts and random lists persists also when the phonological layer is not included in the analysis. This is an effect arising from the nonlinear combination of the shortest paths in the multiplex structure. While on individual layers (free associations and synonyms), there is no statistical difference when they are considered together with generalisations, the resulting multiplex representation displays a higher closeness for words in cohorts rather than for randomised lists (sign test, $n_+ = 71$, $p = 10^{-5}$). Importantly, this difference is not due to generalisations. A difference in closeness between cohorts and random expectation arises also in the multiplex network having only free associations and synonyms as network layers (sign test, $n_+ = 69$, $p = 0.001$). Although individual layers do not display indications of cohort priming, the multiplex lexical network structure does. Since in the model all layers except the phonological one represent semantic memory, this finding is an indication that cohort priming is not exclusively due to phonology but is present also in the combined semantic aspects of the English language.

When network links are rewired at random in configuration models (see Methods), differences in closeness vanish on all individual layers. Figure 3(b) reports the average closeness of empirical cohorts on the randomised network structure. The sign tests give the following results for the layers: $n_+ = 52$, $p = 0.69$ for free associations, $n_+ = 55$, $p = 0.31$ for synonyms, $n_+ = 53$, $p = 0.55$ for generalisations, and $n_+ = 60$, $p = 0.05$ for phonological similarities. The above results indicate that the cognitive advantage expressed by closeness [14, 15, 25] depends on the global structure of individual layers and not on the heterogeneity in the allocation of similarities words might have on each layer when considered individually (e.g., heterogeneity on phonological neighbourhood sizes on the phonological layer, number of associates to a word, etc.). However, even in the configuration models, words in cohorts are closer than random expectation on the whole multiplex structure (sign test, $n_+ = 63$, $p = 0.009$). On the whole multiplex structure, the degree heterogeneity of individual layers gets combined together, so that preserving

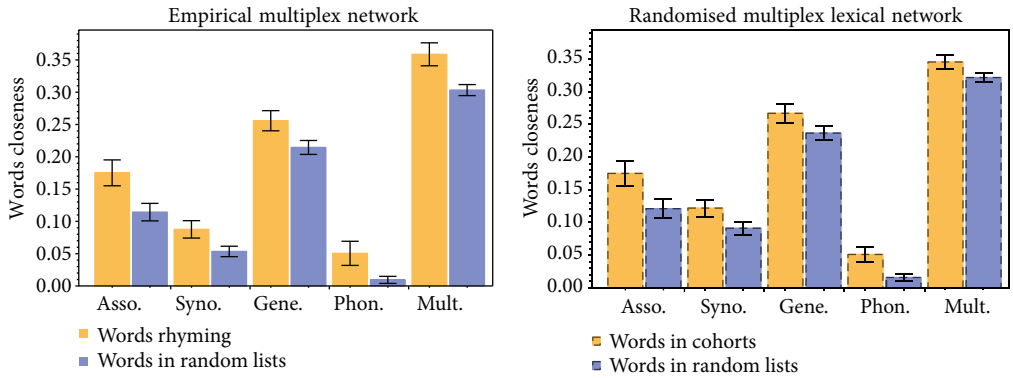


FIGURE 4: (a) Mean closeness distance of words either in cohorts or in randomised lists in the layers, respectively, made of free associations (Asso.), synonyms (Syno.), generalisations (Gene.), and phonological similarities (Phon.). (b) The same as in (a) but for randomised layers of word-word similarities.

degree correlations across layers ultimately still leads to traces of cohort priming effects. This finding indicates that degree heterogeneity determines the availability of shortcuts among words in the same cohort. It also further indicates that priming emerges from the multiplex combination of different aspects of language.

3.3. The Multiplex Lexical Network Identifies Rhyme Priming. As with cohorts, comparisons of the average closeness of words in a rhyming class against one of the words from randomised lists are performed (see Methods).

Figure 4(a) compares the median closeness of words in rhyme classes (orange bars) and in random lists (blue bars) on individual layers and over the whole multiplex structure. Error bars indicate error margins over the median. At the significance level $\alpha = 0.05$, the differences in closeness between words in rhyme classes and words in null models are not statistically significant only on the free association layer (sign test, $n_+ = 30$, $p = 0.26$). Statistically significant differences are observed on the synonyms layer (sign test, $n_+ = 42$, $p < 10^{-5}$), the generalisations layer (sign test, $n_+ = 36$, $p = 0.005$), and the phonological layer (sign test, $n_+ = 41$, $p = 10^{-5}$). A statistically significant difference is found also on the whole multiplex structure (sign test, $n_+ = 42$, $p < 10^{-5}$).

Analogous to cohorts, words in rhyme classes are on average closer than random expectation, indicating a cognitive advantage [14, 15, 25] in processing them together and hence a priming effect. More in detail, this structure suggests a cognitive advantage in lexical processing, assuming cognition is driven by similar network structures and assumptions based on lexical similarity. The magnitude of the gaps in closeness between rhyme classes and random expectations do not correlate with class size (Kendall Tau $|\tau| < 0.06$, p values > 0.5).

Interestingly, rhyme priming persists on one layer more than cohort priming. The layer of synonyms does not display cohort priming but features rhyme priming instead. Notice that rhyme priming persists also in the structure of semantic memory represented by free associations, synonyms, and generalisations (sign test, $n_+ = 37$, $p = 0.002$), again indicating that the multiplex interplay between individual aspects of language can provide evidence of priming effects that might be partially absent when these aspects are considered separately.

When network links are rewired at random in configuration models (see Methods), differences in closeness vanish on all individual layers. Figure 3(b) reports the average closeness of empirical cohorts on the randomised network structure. The sign tests give the following results for the layers: $n_+ = 32$, $p = 0.09$ for free associations, $n_+ = 32$, $p = 0.09$ for synonyms, $n_+ = 29$, $p = 0.40$ for generalisations, and $n_+ = 33$, $p = 0.05$ for phonological similarities. Even in the configuration models, words in rhyme classes are closer than random expectation on the whole multiplex structure (sign test, $n_+ = 35$, $p = 0.01$). Analogous to what happens with cohorts, this result indicates that degree heterogeneities of individual layers get combined together and provide shortcuts to rhyming words that still relate to rhyme priming effects. It has to be underlined that in configuration models rewiring is random but it is always constrained by degree, so that some core-periphery structure induced on the network by the degree distribution can still be present even under randomisation of links. Here, random rewiring does not disrupt shortcuts among words in the same rhyme-class. This indicates that the degree of heterogeneity in the allocation of word-word similarities and the multiplex combination of layers are both important factors for determining rhyme (and cohort) priming.

Notice that the closeness of words is lower in the phonological network compared to other network layers for both cohorts and rhyme classes. This indicates that words in cohorts/rhyme classes tend to cluster more on semantic layers rather than on the phonological layer, even though the considered groups of words are relative to phonological priming. This difference is compatible with the fact that the phonological layer includes words with an average of six phonemes, so that even words sharing on average three phonemes in their onsets or at their end might not have edit distance equal to one and hence they might not be connected with each other. Furthermore, the phonological layer is

significantly more disconnected than the other semantic layers (cf. [9, 22]), so that the lower closeness might be due to words being in different connected components of the phonological network. Notice that if word clustering was a consequence of the definition of the phonological layer, then also randomly selected words should be clustering to a similar extent when compared to words in cohorts and rhyme classes. Instead, the presence of a closeness gap on the phonological layer indicates that words in cohorts and rhyme classes tend to belong to the same connected component of the phonological layer.

4. Discussion

Through the framework of multiplex lexical networks, this paper provides an elegant model to account for and predict potential cognitive advantages [14, 15, 25] in processing together words sharing the same onset or rhyme together. Comparison against null models indicates that these priming effects can be detected already, but not exclusively, at the structural level of word-word similarities when multiple sources of linguistic relations are integrated together rather than indirectly measured with latencies in a laboratory task. The results reported in this analysis correspond to previous work on priming in the psycholinguistic literature and open novel modelling challenges in the investigation of priming through complex networks.

First, the persistence of phonological priming patterns also outside of the phonological layer is an additional confirmation of a nonarbitrariness of language in terms of form-to-meaning correspondences [5, 62] (e.g., English words sharing the onset "sn-" expressing mainly concepts related to "nose"). For a given language, nonarbitrariness refers to the existence of statistical relationships between sound patterns and semantic usage of classes of words. This systematicity corresponds to facilitatory effects in terms of early word learning [62], i.e., children learning words more accurately when spotting systematic and language-specific relations between form and semantic category. The result of phonological priming effects arising also from the combination of hundreds of thousands of semantic, multiplex word-word similarities provides quantitative and large-scale evidence of a *nonrandom* semantic organisation of language that is influenced by phonological regularities such as onset sharing or rhyming.

It is important to underline that cohort and rhyme priming effects have long been detected and investigated in experimental psychology [31, 33, 38, 49], although evidence for them was based only on small samples of hundreds of words being tested in memory-related tasks. The novelty of the current approach is that it is directly based on the large-scale structure of hundreds of thousands of word-word similarities among thousands of commonly used English words interrelated across several semantic and phonological aspects of language. The current network approach is therefore different from an experimental setup from psycholinguistics; in that the network paradigm scales up and tests thousands of words in a considerably easier way compared to the time and effort required in working with subjects in experiments. Also, network representations rely on experiments, but once built, a network can then be used for testing a wide variety of conjectures. For instance, the same network of free associations has been used multiple times for detecting patterns of word learning [7–9], identifying individual creativity levels [2, 3, 11], or even predicting word production in clinical populations [16]. The increasing adoption of complex network models in the cognitive sciences can be beneficial in terms of quantifying large-scale patterns of language usage and acquisition, mainly because of the high versatility of network models [4, 6, 11, 17, 55]. It must also be underlined that network representations bear some assumptions with them and are indeed approximated representations of complex systems. For instance, the multiplex lexical network assumes that all links are weighted equally and are always present over time but this might not be the case in a structure as dynamic as the mental lexicon [5]. Understanding to which extent a network approach is valuable always requires comparison with empirical evidence, often provided by smaller-scale experimental studies. A synergy between theoretical network models and experimental psycholinguistic data represents a valuable combination for future cutting-edge research, a possibility made more appealing by the recent availability of larger digital corpora and massive online psycholinguistic datasets like Opensubtitles [52].

Network approaches must work in synergy with experimental data and more specific experimental setup in order to answer the challenges revealed by network structure. An important example is the attribution of a facilitatory or inhibitory nature to the closeness gaps identified in the current investigation. In fact, a shorter distance among words in cohorts or rhyme classes could also mean higher competition levels among words and hence have an inhibitory, rather than facilitatory, effect on word processing [28, 38, 46]. However, previous experimental studies found that cohort competition effects are stronger for larger cohorts [28, 38]: the more words are activated the stronger the competition effect. This competition is present at phonological and also at semantic levels, and it leads to slower performance on lexical decision tasks. In the current investigation, both smaller (i.e., comprising 20 words) and larger (i.e., comprising 100 words) cohorts consistently displayed the same priming patterns reported in the manuscript. Differently put, words in cohorts are always closer than random expectation on the multiplex lexical structure and this gap is independent of cohort size. Since competition effects are size-dependent [28, 47] while priming effects are not [46], this finding might be an important indication that the differences in the shortest path lengths found in this work represent mainly priming effects rather than lexical competition. Assessing the facilitatory or inhibitory nature of these priming patterns requires additional empirical data and represents an interesting future research direction.

Notice that cohort priming is not the only effect driving lexical retrieval. The cohort model neglects important aspects of language such as syntactic structure, which can significantly alter access to semantic memory [3, 11, 33]. Recently, experiments from cognitive neuroscience have indicated that cohort effects and lexical competition levels are present when words are processed individually while competition is absent

when words are heard in short sentences [28]. Also, semantic information aided the discrimination process of words in larger cohorts [47]. The disappearance of cohort competition effects in sentences or in presence of semantic information indicates that word similarities and syntactic structure are both highly important in driving activation to specific target words, thus ultimately having a facilitatory, rather than inhibitory, effect on lexical retrieval. Although not fully coincident with the same richness of semantic information from sentences, the adopted multiplex representation does not consider words as disconnected units but rather provides information also about word context through similarities, e.g., the link between "play" and "act" represents the context of theatrical plays and the link between "play" and "football" represents the context of games. Hence, previous findings of context [28] and semantic word similarities [47] reducing lexico-phonological competition might represent an additional indication that the patterns found in this investigation are facilitatory rather than inhibitory. Notice also that in the relevant literature there is strong evidence for facilitatory priming to correlate positively with concept relatedness [35–37] and for inhibitory priming to be mainly driven by ignoring unrelated concepts [5, 35, 61]. Combining this literature with the recent studies indicating that shorter network distance is a valid proxy for closer conceptual relatedness [15, 57] further indicates that the priming effects detected on the multiplex structure are mainly facilitatory. This is in agreement also with the previous experiments specifically focused on phonological priming and indicating that cohort effects facilitate word memorisation [31, 33, 46] while rhyming facilitates phonological awareness, specifically in children [38]. In order to fully address the nature of the patterns highlighted by the multiplex structure, a psycholinguistic experiment involving the cohort/rhyming words analysed in this investigation would be an important future research direction. By considering reaction times in a lexical decision task, it would be interesting to understand if there is any critical threshold c^* of closeness above which lexical competition might overcome facilitation, e.g., lexical items being so close that they can be confused, thus inhibiting retrieval of the correct item. Another interesting research direction would be correlating closeness gaps to competition effects in cohort priming arising by interactions of specific word suffixes, which can inhibit one another [63].

From a network perspective, the current investigation provides additional empirical evidence that multiplex networks can highlight phenomena that cannot be detected by single-layer networks. In fact, for both cohorts and rhyme classes, individual layers do not always display priming effects, while the multiplex network obtained by combining together these layers always highlighted statistically significant differences in terms of network distances. By assuming that these differences indicate a cognitive facilitation in processing words together, as indicated by many recent studies [15, 25, 57], then the above results quantitatively indicate that cohort and rhyme priming can arise from an interplay between either different aspects of semantic memory (e.g., synonyms and free associations) or by an interplay between different aspects of whole mental lexicon (e.g., phonological similarities and free associations). More in detail, assuming that lexical retrieval is influenced by a multilayer network structure of the mental lexicon, phonological priming effects might then be an emergent property of the adopted multiplex representation of the lexicon as it arises from the multiple interactions among words across different aspects of language.

Notice that the gap in closeness between cohorts/rhyme classes and random expectation in the empirical multiplex network is almost an order of magnitude larger than in the randomly rewired multiplex network, containing random links between words. This indicates that the detected gaps in closeness are mainly due to the empirical structure of word-word similarities in the real layers rather than to the act of combining layers, instead. Notice also that the current investigation cannot provide any causality link, since the structure itself is unable to fully identify the nature of the priming patterns found in the literature, as these patterns are heavily influenced by other aspects of lexical retrieval such as attention [35], modality [28, 36], and timing between prime and target [37]. Addressing through experiments the challenges opened by the current multiplex network investigation on priming would also require a more thorough investigation of the factors influencing priming beyond the mental lexicon structure, such as different stimulus onset asynchrony determining the strength of positive priming [35, 37] or different modalities affecting the extent of negative semantic priming [35]. This rich variety of priming patterns underlines the importance of further multilayer modelling efforts for the understanding of priming effects in language-related tasks.

One limitation of the original cohort model was that it neglected the influence that semantics exerts over lexical retrieval in perceptual tasks [31, 41], an element that is taken into account in more refined models of word processing [41, 43] and confirmed also by experimental studies [28]. Interestingly, the fact that free associations displayed a significant gap in closeness for rhyming but not for cohorts might be a consequence of the different positions of phonemes. Relying on the last phonemes would allow for a temporal unfolding to occur, during which the first part of the word would be acquired and some of its semantic features would be available for processing, features that cannot be available when the first phonemes are heard instead. This difference reconciles the finding that in rhyme priming there is a closeness gap also in free associations, a gap that is absent when cohorts are considered. This quantitative difference indicates that rhyme priming is more heavily influenced by semantic information compared to cohort priming.

A limitation of the multiplex approach is that it does not consider individual variability. It is expected for lexical retrieval to be influenced also by individual factors such as fluid intelligence or other active cognitive search strategies [28, 42, 43]. Even creativity levels have been recently shown to deeply influence lexical retrieval and word identification in healthy populations [3, 11, 55]. One possibility for overcoming this limitation could be the substitution of the layer of free associations with other empirical layers, always of free associations but obtained from subjects

belonging to a specific population, like for instance highly creative people. Previous research has shown that more creative people tend to associate even semantically unrelated concepts [3, 11, 55], so that new shortcuts might appear in the free association layer. These paths might alter the results found in the current investigation for normative subjects. Considering other ad hoc layers of free associations could also be a valuable research direction for generalising the model in order to incorporate ageing. Recent work has shown that over time the mental lexicon undergoes some substantial changes and some word-word similarities get lost [19], thus potentially altering the shortcuts connecting words in cohorts or rhyme classes. A reduction of priming effects with age is expected, particularly the one due to rhyming which has been empirically shown to decrease in strength from childhood to adulthood [38].

Also, the investigation of clinical populations could be interesting for future research [17, 24]. In case the shortcuts allowing for cohort and rhyme classes were resilient to progressive word failure in people with aphasia, these word-word associations might be used for designing strategies of intervention for restoring or mending the functionality of the mental lexicon. The framework of multiplex lexical networks has been already applied to clinical populations with aphasia [16], and it showed that word production in subjects with aphasia crucially depends on the closeness that words have over the multiplex lexical structure. Words with higher closeness centrality were easier to pronounce in picture naming tasks compared to words with lower closeness. Investigating potential differences between words in cohorts/rhyme classes and specific null models would represent an interesting research direction.

All in all, multiplex lexical networks represent a powerful framework for the quantitative investigation of psycholinguistic patterns where the interplay between different semantic and phonological aspects of language is relevant. The multiplex structure of these linguistic networks opens new important challenges for the large-scale understanding of the cognitive processes driving language usage.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The author acknowledges Nichol Castro for the insightful discussion and constructive feedback about both the conceptual and style aspects of this manuscript. The author also acknowledges Markus Brede and Manlio De Domenico for stimulating discussions in the early stages of this work. M.S. was supported by the EPSRC Doctoral Training Centre grant (EP/G03690X/1).

References

[1] A. Baronchelli, R. Ferrer-i-Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen, "Networks in cognitive science," *Trends in Cognitive Sciences*, vol. 17, no. 7, pp. 348–360, 2013.

[2] S. De Deyne, Y. N. Kenett, D. Anaki, M. Faust, and D. Navarro, "Large-scale network representations of semantics in the mental lexicon," in *Big Data in Cognitive Science*, M. N. Jones, Ed., pp. 174–202, Psychology Press: Taylor & Francis, 2017.

[3] Y. N. Kenett, O. Levy, D. Y. Kenett, H. E. Stanley, M. Faust, and S. Havlin, "Flexibility of thought in high creative individuals represented by percolation analysis," *Proceedings of the National Academy of Sciences*, vol. 115, no. 5, pp. 867–872, 2018.

[4] M. S. Vitevitch, "What can graph theory tell us about word learning and lexical retrieval?," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 2, pp. 408–422, 2008.

[5] J. Aitchison, *Words in the Mind: An Introduction to the Mental Lexicon*, John Wiley & Sons, 2012.

[6] N. Beckage, L. Smith, and T. Hills, "Small worlds and semantic network growth in typical and late talkers," *PLoS One*, vol. 6, no. 5, article e19348, 2011.

[7] A. E. Sizemore, E. A. Karuza, C. Giusti, and D. S. Bassett, "Knowledge gaps in the early growth of semantic networks," 2017, http://arxiv.org/abs/1709.00133.

[8] M. Stella, N. M. Beckage, and M. Brede, "Multiplex lexical networks reveal patterns in early word acquisition in children," *Scientific Reports*, vol. 7, no. 1, article 46730, 2017.

[9] M. Stella, N. M. Beckage, M. Brede, and M. De Domenico, "Multiplex model of mental lexicon reveals explosive learning in humans," *Scientific Reports*, vol. 8, no. 1, article 2259, 2018.

[10] M. Stella and M. De Domenico, "Distance entropy cartography characterises centrality in complex networks," *Entropy*, vol. 20, no. 4, p. 268, 2018.

[11] Y. N. Kenett, "13 - going the extra creative mile: the role of semantic distance in creativity – theory, research, and measurement," in *The Cambridge Handbook of the Neuroscience of Creativity*, R. Jung and O. Vartanian, Eds., pp. 233–248, Cambridge University Press, Cambridge, 2018.

[12] C. S. Q. Siew and M. S. Vitevitch, "Spoken word recognition and serial recall of words from components in the phonological network," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 42, no. 3, pp. 394–410, 2016.

[13] M. S. Vitevitch, K. Y. Chan, and R. Goldstein, "Insights into failed lexical retrieval from network science," *Cognitive Psychology*, vol. 68, pp. 1–32, 2014.

[14] M. S. Vitevitch and P. A. Luce, "Phonological neighborhood effects in spoken word perception and production," *Annual Review of Linguistics*, vol. 2, no. 1, pp. 75–94, 2016.

[15] Y. N. Kenett, E. Levi, D. Anaki, and M. Faust, "The semantic distance task: quantifying semantic distance with semantic network path length," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 43, no. 9, pp. 1470–1489, 2017.

[16] N. Castro and M. Stella, "The multiplex structure of the mental lexicon influences picture naming in people with aphasia," PsyArXiv, 2018.

[17] M. S. Vitevitch and N. Castro, "Using network science in the language sciences and clinic," *International Journal of Speechlanguage Pathology*, vol. 17, no. 1, pp. 13–25, 2015.

[18] M. S. Vitevitch, R. Goldstein, C. S. Siew, and N. Castro, "Using complex networks to understand the mental lexicon," in *Yearbook of the Poznan Linguistic Meeting, Vol. 1*, pp. 119–138, De Gruyter Open, 2014.

[19] D. M. Burke and M. A. Shafto, "Language and aging," in *The Handbook of Aging and Cognition*, F. I. M. Craik and T. A. Salthouse, Eds., pp. 373–443, Psychology Press, New York, NY, USA, 2008.

[20] P. A. Luce and D. B. Pisoni, "Recognizing spoken words: the neighborhood activation model," *Ear and Hearing*, vol. 19, no. 1, pp. 1–36, 1998.

[21] M. S. Vitevitch and M. S. Sommers, "The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults," *Memory & Cognition*, vol. 31, no. 4, pp. 491–504, 2003.

[22] M. Stella and M. Brede, "Patterns in the English language: phonological networks, percolation and assembly models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 5, article P05006, 2015.

[23] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Longitudinal analysis of early semantic networks," *Psychological Science*, vol. 20, no. 6, pp. 729–739, 2009.

[24] J. Borge-Holthoefer, Y. Moreno, and A. Arenas, "Modeling abnormal priming in Alzheimer's patients with a free association network," *PLoS One*, vol. 6, no. 8, article e22651, 2011.

[25] R. Goldstein and M. S. Vitevitch, "The influence of closeness centrality on lexical processing," *Frontiers in Psychology*, vol. 8, p. 1683, 2017.

[26] D. Fay and A. Cutler, "Malapropisms and the structure of the mental lexicon," *Linguistic Inquiry*, vol. 8, pp. 505–520, 1977.

[27] A. Pisoni, M. Cerciello, Z. Cattaneo, and C. Papagno, "Phonological facilitation in picture naming: when and where? A tDCS study," *Neuroscience*, vol. 352, pp. 106–121, 2017.

[28] J. Zhuang and B. J. Devereux, "Phonological and syntactic competition effects in spoken word recognition: evidence from corpus-based statistics," *Language, Cognition and Neuroscience*, vol. 32, no. 2, pp. 221–235, 2017.

[29] I. Dautriche, D. Swingley, and A. Christophe, *Cognition*, vol. 143, pp. 77–86, 2015.

[30] N. E. Evangelopoulos, "Latent semantic analysis," vol. 4, no. 6, pp. 683–692, 2013.

[31] W. D. Marslen-Wilson, "Functional parallelism in spoken word-recognition," *Cognition*, vol. 25, no. 1-2, pp. 71–102, 1987.

[32] D. E. Meyer and R. W. Schvaneveldt, "Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations," *Journal of Experimental Psychology*, vol. 90, no. 2, pp. 227–234, 1971.

[33] M. Taft and G. Hambly, "Exploring the cohort model of spoken word recognition," *Cognition*, vol. 22, no. 3, pp. 259–282, 1986.

[34] T. K. Landauer, "Latent semantic analysis," Wiley Online Library, 2006.

[35] E. Weingarten, Q. Chen, M. McAdams, J. Yi, J. Hepler, and D. Albarracín, "From primed concepts to action: a meta-analysis of the behavioral effects of incidentally presented words," *Psychological Bulletin*, vol. 142, no. 5, pp. 472–497, 2016.

[36] D. J. Foss, "A discourse on semantic priming," *Cognitive Psychology*, vol. 14, no. 4, pp. 590–607, 1982.

[37] A. M. B. de Groot, A. J. W. M. Thomassen, and P. T. W. Hudson, "Primed-lexical decision: the effect of varying the stimulus-onset asynchrony of prime and target," *Acta Psychologica*, vol. 61, no. 1, pp. 17–36, 1986.

[38] P. J. Brooks and B. MacWhinney, "Phonological priming in children's picture naming," *Journal of Child Language*, vol. 27, no. 2, pp. 335–366, 2000.

[39] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing," *Psychological Review*, vol. 82, no. 6, pp. 407–428, 1975.

[40] A. M. Collins and M. R. Quillian, "Retrieval time from semantic memory 1," *Journal of Verbal Learning and Verbal Behavior*, vol. 8, no. 2, pp. 240–247, 1969.

[41] G. S. Dell, "A spreading-activation theory of retrieval in sentence production," *Psychological Review*, vol. 93, no. 3, pp. 283–321, 1986.

[42] M. G. Gaskell and W. D. Marslen-Wilson, "Integrating form and meaning: a distributed model of speech perception," *Language and Cognitive Processes*, vol. 12, no. 5-6, pp. 613–656, 1997.

[43] A. Lahiri and W. Marslen-Wilson, "The mental representation of lexical form: a phonological approach to the recognition lexicon," *Cognition*, vol. 38, no. 3, pp. 245–294, 1991.

[44] M. S. Vitevitch, K. Y. Chan, and S. Roodenrys, "Complex network structure influences processing in long-term and short-term memory," *Journal of Memory and Language*, vol. 67, no. 1, pp. 30–44, 2012.

[45] M. Stella and M. BredeH. Cherifi, B. Gonçalves, R. Menezes, and R. Sinatra, "Mental lexicon growth modelling reveals the multiplexity of the English language," in *Complex Networks VII. Studies in Computational Intelligence*, pp. 267–279, Springer, Cham, 2016.

[46] L. M. Slowiaczek, H. C. Nusbaum, and D. B. Pisono, "Phonological priming in auditory word recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 13, no. 1, pp. 64–75, 1987.

[47] J. Zhuang, B. Randall, E. A. Stamatakis, W. D. Marslen-Wilson, and L. K. Tyler, *Journal of Cognitive Neuroscience*, vol. 23, no. 12, pp. 3778–3790, 2011.

[48] J. K. Gordon and S. R. Baum, "Rhyme priming in aphasia: the role of phonology in lexical access," *Brain and Language*, vol. 47, no. 4, pp. 661–683, 1994.

[49] D. N. Rapp and A. G. Samuel, "A reason to rhyme: phonological and semantic influences on lexical access," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 28, no. 3, pp. 564–571, 2002.

[50] M. Coltheart, "The MRC psycholinguistic database," *The Quarterly Journal of Experimental Psychology*, vol. 33, no. 4, pp. 497–505, 1981.

[51] G. A. Miller, "WordNet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[52] A. Barbaresi, *Language-classified Open Subtitles (LACLOS): download, extraction, and quality assessment*, Ph.D. thesis, BBAW, 2014.

[53] F. Battiston, V. Nicosia, and V. Latora, "The new challenges of multiplex networks: measures and models," *The European Physical Journal Special Topics*, vol. 226, no. 3, pp. 401–416, 2017.

[54] M. De Domenico, A. Solè-Ribalta, E. Cozzo et al., "Mathematical formulation of multilayer networks," *Physical Review X*, vol. 3, article 041022, 2013.

[55] Y. N. Kenett, R. E. Beaty, P. J. Silvia, D. Anaki, and M. Faust, "Structure and flexibility: investigating the relation between the structure of the mental lexicon, fluid intelligence, and creative achievement," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 10, no. 4, pp. 377–388, 2016.

[56] "RhymeZone dictionary," March 2018, https://www.rhymezone.com/.

[57] M. S. Vitevitch, R. Goldstein, and E. Johnson, "Path-length and the misperception of speech: insights from Network Science and Psycholinguistics," in *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, A. Mehler, A. Lücking, S. Banisch, P. Blanchard, and B. Job, Eds., pp. 29–45, Springer, Berlin, Heidelberg, 2016.

[58] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, article 198701, 2001.

[59] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures & Algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.

[60] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Communications*, vol. 6, no. 1, p. 6864, 2015.

[61] G. Houghton and S. P. Tipper, "A model of inhibitory mechanisms in selective attention," Academic Press Ltd., 1984.

[62] M. Dingemanse, D. E. Blasi, G. Lupyan, M. H. Christiansen, and P. Monaghan, "Arbitrariness, iconicity, and systematicity in language," *Trends in Cognitive Sciences*, vol. 19, no. 10, pp. 603–615, 2015.

[63] J. C. L. Ingram, *Neurolinguistics: An Introduction to Spoken Language Processing and Its Disorders*, Cambridge University Press, 2007.