The Effects of Time-Averaging on Archaeological Networks

Dries Daems (ORCID: 0000-0002-6444-9013)[1,*], Emily Coco (ORCID: 0000-0002-9200-8469)[2], Andrew Gillreath-Brown (ORCID: 0000-0002-2272-5542)[3,4,5], Danai Kafetzaki (ORCID: 0000-0001-7002-5850)[1,6]
[1] Sagalassos Archaeological Research Project, KU Leuven
[2] Center for the Study of Human Origins, Department of Anthropology, New York University
[3] Scripps Institution of Oceanography, University of California San Diego
[4] Department of Anthropology, Washington State University
[5] Max Planck Institute for Geoanthropology, Department of Archaeology
[6] Center for Statistics, Data Science Institute, Hasselt University
* Corresponding author: dries.daems@kuleuven.be

Abstract

It is well recognized that time-averaging of archaeological deposits results in significant biases in interpretations of the archaeological record. In this study, we investigate the biases introduced by time-averaging in the study of social and economic networks from the archaeological record. Using three different archaeological network datasets, we combine network slices from multiple time periods to mimic the effects of time-averaging to understand how the palimpsest nature of the archaeological record affects our interpretations of the network. The results of our analysis indicate that time-averaging reduces the fidelity of network interpretations compared to the non-time-averaged networks when analyzing network or node properties. Our results also showed that the effects of time-averaging are highly dependent on initial network structures. This makes it difficult to establish general rules for how to interpret time-averaged networks in archaeology. However, our study shows that it is of paramount importance that archaeologists are aware of these biases and evaluate the reliability of their data accordingly.

Keywords

archaeology; network science; analytical biases; archaeological networks

Introduction

The concept of time is at the core of archaeology. The discipline is built around the fundamental act of chronometry (Holdaway and Wandsnider 2008). Archaeology studies the activities of people, groups, and societies and how they changed through time. Methods such as seriation and time series analysis provide quantitative tools to address temporal changes across different scales. For this, we rely on data with distinct time signatures. Unfortunately, the

temporal resolution of archaeological data is often coarse, impeding fine-grained analysis. The archaeological record is commonly described as a palimpsest, that is, characterized by the loss of temporal resolution through the compression of multiple diachronic levels into a single, synchronic layer. This effect has also been described as time-averaging, the association of materials produced by diachronic events into the same unit, either through depositional processes, disturbance factors, or coarse chronological resolution of archaeological interpretation (Perreault 2018). Time-averaging results in significant bias and information loss in archaeological interpretation (see Contreras and Meadows 2014; Miller-Atkins and Premo 2018; Perreault 2018; 2019; Surovell et al. 2009; Surovell and Brantingham 2007). Yet, the effects of time-averaging on archaeological data are rarely accounted for in our analyses and interpretations.

In this paper, we address some of the fundamental methodological implications of time-averaging in preparing and interpreting archaeological data for network analysis. Temporal analysis of archaeological networks (e.g., networks constructed from material culture and archaeological sites) requires the creation of time slices that can be analyzed separately or mashed together to simulate time-averaging effects. We use a time-averaging workflow to combine network edge lists from multiple time slices (from a minimum of two up to the total amount of time slices defined by the temporal scope of the dataset divided by the bin size) to create a new “time-averaged” network. Network topologies from the time-averaged networks are then compared to the “non-time-averaged” networks - hereafter designated as “original” networks - to determine how time-averaging affects the network analysis and subsequent archaeological interpretation of the network. The values of the time-averaged networks are compared to a null model of random networks to determine whether time-averaging shows distinct properties or results in seemingly random network topologies.

We apply these methodological steps to archaeological datasets of material distributions and settlement and road networks. First, we use the Prignano and colleagues (2019) road and settlement network dataset, where they examined inter-polity interactions in southern Etruria from 950–500 BCE. Second, we use Mills and colleagues (2018) apportioned ceramic data and associated similarity matrices constructed from the Chaco Social Networks database (Mills et al. 2018; Peeples et al. 2016), in which they evaluated how migration contributed to the creation of the “Chaco World” (800–1200 CE).Third, we use pottery distribution data from the ICRATES dataset (Bes 2015). The ICRATES database contains 33,939 diagnostic tableware sherds dated to the Hellenistic and Roman periods (300 BCE–700 CE) from 275 sites throughout the Eastern Mediterranean. We purposefully chose three network datasets made with different types of archaeological data and derived from different temporal and spatial contexts to maximize the generalization of our findings for the archaeological discipline.

The results of our analysis confirm that, for most metrics, time-averaging reduces the fidelity of network interpretations compared to the original networks when analyzing network or node properties. Time-averaged networks are typically larger in size, and have higher average betweenness centrality, path length, and diameter. In the interpretation of archaeological networks, this could lead archaeologists to overestimate the extent of settlement hierarchies, the importance of main hubs in political and economic networks, or the degree of integration and efficiency of information flows within the network. Our analysis also demonstrates that

increased time-averaging decreases the similarity of node sets identified as having high centrality values. This is important given that archaeologists frequently use centrality measures of nodes to determine the relative (social, political, and economic) positions of their data points, for example settlements within a regional network. All of our results indicate that archaeological interpretations of network properties and positions from the archaeological record are biased due to the time-averaged nature of the archaeological record. It is of paramount importance that archaeologists are aware of these biases and evaluate the reliability of their data accordingly.

Materials and methods

Data sets

For this study, we used three archaeological datasets consisting of artifact distributions and settlement-road networks to create time-averaged networks. The first, a settlement and road network dataset, was developed by Prignano and colleagues (2019) to understand inter-polity interactions in southern Etruria from 950 to 500 BCE. Their study used road maps to create route networks between settlements in five different time slices; networks were dated to the Early Iron Age 1 Early (950–900 BCE), the Early Iron Age 1 Late (900–825 BCE), the Early Iron Age 2 (850–720 BCE), the Orientalizing Age (730–580 BCE), and the Archaic Period (580–500 BCE).

The second dataset consists of ceramic data from the Chaco Social Networks database (Peeples et al. 2016). The Chaco Social Networks database contains more than 1.9 million sherds that fit into 28 base ceramic types (or 352 refined types) from 280 sites (based on the Southwest Social Networks project (SWSN) ID number) throughout the southwestern United States (Peeples et al. 2016). We use apportioned ceramic data and associated similarity matrices from Mills and colleagues (2018). The ceramic assemblages were apportioned into 50-year intervals, but here we use the 25-year interval data, which cover the Late Pueblo I (PI) period through the end of the Pueblo III (PIII) period. These data are available at https://cybersw.org/ (Mills et al. 2020).

The third dataset consists of pottery distribution data from the ICRATES dataset (Bes 2015; Bes et al. 2019). The ICRATES database contains nearly 34,000 diagnostic tableware sherds dated to the Hellenistic and Roman periods (300 BCE–700 CE) from 275 sites throughout the Eastern Mediterranean (Bes et al. 2019). The most important wares included in the dataset are: Eastern Sigillata A (ESA), Eastern Sigillata B (ESB), Eastern Sigillata C (ESC), Eastern Sigillata D (ESD), Italian Sigillata (ITS), African Red Slip Ware (ARSW), Cypriot Red Slip Ware (CRSW), and Phocaean Red Slip Ware (PRSW). We used the ICRATES dataset openly available on the Archaeology Data Service repository (https://archaeologydataservice.ac.uk/archives/view/icrates_lt_2018/).

The three selected datasets cover multiple types of archaeological data (settlements, roads and material culture), scales, resolutions and dimensions, going from relatively small, low density, and regional-level (Etrurian roads and settlements), over mid-large, high density and inter-regional (ICRATES), to large-scale, mid-density, and inter-regional networks (Chaco). This

way, we are able to show the effects of time-averaging across a range of scales relevant for archaeological research. Furthermore, the three datasets allow us to create different types of networks: a road network, a co-presence network, and a similarity network. This allows us to understand if any of these types of networks are more or less robust to the effects of time-averaging.

Network creation

The settlement-road network for southern Etruria was already available as five timeslices, each consisting of a set of nodes and edges per time period. The ICRATES and Chaco datasets, on the other hand, only contained a list of locations and concordant pottery data catalogs. For these two datasets, we created edge lists based on shared wares between sites and time slices on the basis of pottery ware chronologies.

The Chaco social networks were created via adjacency matrices where adjacency between two given sites was based on similarity of pottery ware frequencies, following the methodology used by Mills and colleagues (Mills et al. 2018 and references therein). Edges were weighted by the Brainard-Robinson similarity value between two sites. The Chaco data was already apportioned into 20 time slices of 25-year intervals, so we maintained those time slices when creating the networks.

For the ICRATES database, we merged the location and pottery data catalog files to create an integrated overview of all relevant information, including locations, chronology and pottery form. We then transformed this data table into an adjacency matrix, in which two or more locations that feature the same pottery form are represented as sharing a link. Fifty time slices were created by defining a fixed time interval of 20 years as bin size for the network partitioning and dividing the full time range of the ICRATES dataset by these intervals. We used a uniform distribution to divide the frequencies per type over the relevant time slices. For example, the counts of a type with a lifespan of 100 years from 200 to 100 BCE would be evenly divided across five time slices. If the lifespan of a type does not fully coincide with an entire time slice, then we apportion the type counts only for that part of the time slice which overlaps with the lifespan of the slice. For example, if a type originated in 170 BCE and disappeared in 150 BCE, we alot half of the counts to the time slices 180-160 BCE and half to the time slice of 160-140 BCE.

It is essential to note an important distinction between the Etruria site-road network on the one hand, and the Chaco and ICRATES networks on the other. The latter two consist of evenly distributed time slices of, respectively, 25 and 20 years, based on pottery chronologies. The former consists of a much coarser temporal resolution based on the dating of sites to specific chronological periods. This also means that individual time slices in the Etruria network are of uneven length (more specifically, 50, 75, 130, 150 and 80 years) based on the duration of each time period. We purposefully chose not to normalize these time slices into an even distribution comparable with Chaco and ICRATES because this kind of uneven temporal distribution is characteristic of many archaeological networks and therefore needs to be included in this analysis as well.

Time-averaging and comparative network analysis

We use a time-averaging workflow to combine network edge lists from multiple time slices, ranging from a minimum of two time slices up to the total amount of time slices (defined by the temporal scope of the dataset divided by the bin size) to create new “time-averaged” networks. For the ICRATES and Etrurian road network datasets, this was accomplished by appending multiple edge lists together and then simplifying the network graph to remove any duplicate edges. This methodology was based on and adapted from earlier work on the effects of time-averaging on ancient food webs to which some of the authors contributed (Shaw et al. 2021). For the Chaco dataset, time-averaging was done by combining lists of ceramic frequency data from multiple time slices and creating new similarity networks, following the strategy used by Mills and colleagues to create 50 year time slices (Mills et al. 2018). Through these workflows, we created 35 networks for the Etrurian dataset, 1540 for Chaco, and 22080 for ICRATES. Network topologies from the time-averaged networks were then calculated and compared to the original network topology to determine how time-averaging affects the network analysis and subsequent archaeological interpretation of the network.

In this study, we look at multiple network metrics that are typically used to describe network topologies and are often used for particular archaeological interpretations (described below): degree centrality, betweenness centrality, eigenvector centrality, clustering coefficient, cohesion, diameter, average shortest path length, and modularity. Centrality measures are used in archaeology to describe social and economic connections because they emphasize the position and importance of individual nodes within the network (Collar et al. 2015; Gjesfjeld 2015; Mills 2017; Mullins 2016). Eigenvector centrality is often used to measure the influence of nodes within the network, leading to its use for identifying hierarchy in settlement systems (Brughmans 2013; Gjesfjeld 2015). Clustering coefficients describe the tendency of nodes to form densely connected groups. Similarly, cohesion or edge density of the network measures the density of ties within the network, and has been used to indicate intentional interconnectedness in archaeological research (Mullins 2016). The diameter of the network measures the shortest path length between the two most distant nodes. Average shortest path length is calculated as the average number of steps on the shortest path between all possible pairs of nodes. This metric is often used to describe the efficiency of information flow on a network. Modularity is often used as a measure of community structure because it reflects the extent to which network edges are concentrated within particular groups. In archaeology, modularity can identify the presence of subgroups or components of the network and is therefore used to address questions of regionalization (Buchanan et al. 2016).

The values of the time-averaged networks are compared to null random networks to determine whether time-averaging results in random network topologies or shows distinct properties. To construct the null model, we rewire all edges of the original graphs 100 times, thus generating a randomized alternative for each original graph. For each of the random graphs, we calculate the network metrics described above to acquire a set of 100 null model values for each metric. We then compare the empirical values of the original and time-averaged networks to the null model values by calculating the model error (ME). The ME is calculated as

the normalized difference between the median null value and the empirical values, for each metric and for each step of the time-averaging process. Normalization is achieved by dividing the aforementioned difference and the denominator depends on the sign of the differences between the median null value and the empirical values. When all differences have a common sign, the normalization is performed with the value of the median null value minus the value at the 5th percentile in case signs are negative, the difference of the value at the 95th percentile and the median null value, in case signs are all positive. When the signs are mixed, the normalization employs the values at the 2.5th percentile and the 97.5th percentile in a similar manner. Values of ME exceeding ±1 are considered significant differences compared to random.

We also compared time-averaged network values to the values for each original network via a principal components analysis (PCA) to understand which network metrics were driving differences between the networks. PCA is a rich display which allows us to interpret the relationship between metrics, plotted as vectors, the relationship between networks, plotted as points, and the relative metric values per network. Metrics with higher variation among networks have higher contribution on the principal components axes and therefore higher length. Metrics which are uncorrelated form an angle of 90 degrees, positively correlated metrics form an angle of zero or a few degrees, while metrics with strong negative correlation form an angle closer to 180 degrees. Similar networks are plotted in closer spatial proximity. The distance from the axis origin, along the projection of the point to the vector, signifies the value for that network with respect to the central value for each metric. In this way, outliers can be spotted. A disadvantage of the biplot is the visual clutter which is created in places; however, we alleviate the clutter by annotating the visualization. It should be noted that some networks are basically duplicated in the output since aggregating network-a and network-b is exactly the same as aggregating network-b and network-a. Also, the color for the points is given based on the ‘latest’ network. Therefore the network resulting from the aggregation of a and b will retain the color of network b. For this analysis, the component space was defined using only the time-averaged network values and then the original values were plotted into that component space.

In addition to network-level measures, we also analyzed how time-averaging affects individual nodes by tracing which nodes are identified as having high or low centrality values. For each original network and each time-averaged network, we determined the set of five nodes with the highest and lowest values of each centrality metric. We then compared via Jaccard similarity how similar the node sets for the time-averaged networks were to the actual node set from the original graph; this indicates whether time-averaging would impact the network structure enough to affect potential archaeological interpretations on both ends of the archaeological spectrum.

All analysis scripts were coded in R (R Core Team 2022). The following R packages were used: igraph (Csardi and Nepusz 2006), rlist (Ren 2016), ggthemes (Arnold 2019), ggraph (Pedersen and RStudio 2021), ggpubr (Kassambara 2020), cowplot (Wilke 2019), gridExtra (Auguie and Antonov 2017), factoextra (Kassambara and Mundt 2019), utils (R Core Team 2022), DT (Xie et al. 2021), rmarkdown (Allaire et al. 2021), and tidyverse (Wickham 2017), which includes dplyr (Wickham et al. 2021), purrr (Henry et al. 2020), ggplot2 (Wickham et al. 2020), stringr (Wickham and RStudio 2019). All scripts are available at https://github.com/cocoemily/time-averaged-networks and https://osf.io/7tpbg.

Sensitivity analysis

As with all archaeological data, it is highly unlikely that our networks fully represent the actual distribution of connections in the past. Various nodes might be missing from the distribution due to a host of reasons such as site destruction, lack of study coverage, or lack of expertise necessary to properly identify the relevant material at hand. Our archaeological networks are therefore only samples of the complete network that—by default—can never be fully recovered. We need to assess how missing nodes might impact our interpretations of the overall network structure and the roles of individual nodes within the network (Costenbader and Valente 2003).

To deal with this uncertainty and assess how our metrics of interest are affected by missing data, we conducted sensitivity analysis (Brughmans and Peeples 2023: chapter 5; Peeples and Brughmans 2023). We used numerous network metrics (more specifically: degree centrality, eigenvector centrality, betweenness, clustering coefficient, and modularity) and evaluated the impacts of node removal by sampling at different percentages across the network (from 90% to 10% in intervals of 10%). We then compared the results of the listed metrics across the sampling fractions to determine how the increasing removal of nodes impacts each of the networks. We followed Matt Peeples's tutorial (2017) for dealing with uncertainties in archaeological data, calculating rank-order correlations (through Spearman's ρ measure) between the metrics for the original network sample, and each of the sub-samples to evaluate the average correlation and error at each sampling fraction.

This allows us to see how robust the network might be for different amounts of time averaging and at different sampling fractions. If a particular network metric is robust in light of the increasing removal of nodes, representing a thorough process of sub-sampling, our confidence in the validity of the metric for our original network increases.

In addition to examining the impact of sub-sampling on network nodes, we also assess the impact of edge removal on the various networks. Here, we follow the same steps as for node removal, but sub-sample for edge removal to see how sensitive the network is to missing edges for the same centrality and network metrics. For both node and edge removal, we performed 100 replicates.

Results

Time-averaging effects on network metrics

Comparison to randomized networks

We first compared the original and time-averaged networks to a null model created by drawing randomized edges from the original graphs to determine whether non-random signals appear as time-averaging increases. Null models allow us to determine if the network topologies

we observe could have been produced by chance. For our particular analysis, we wanted to determine if time-averaging has non-random effects on network topology.

Etruria

For the Etrurian settlement-road network dataset, we see significant differences between the null model values and the time-averaged networks (see Figure 1). The original network values are also clearly distinct from the null model. Additionally, we can see that the trendlines for comparison to the null models are relatively flat, indicating that the values of the network metrics do not change significantly as time-averaging increases. This suggests that some sort of signal from the original network is maintained even as time-averaging increases, making these metrics appear relatively robust to time-averaging for this dataset.

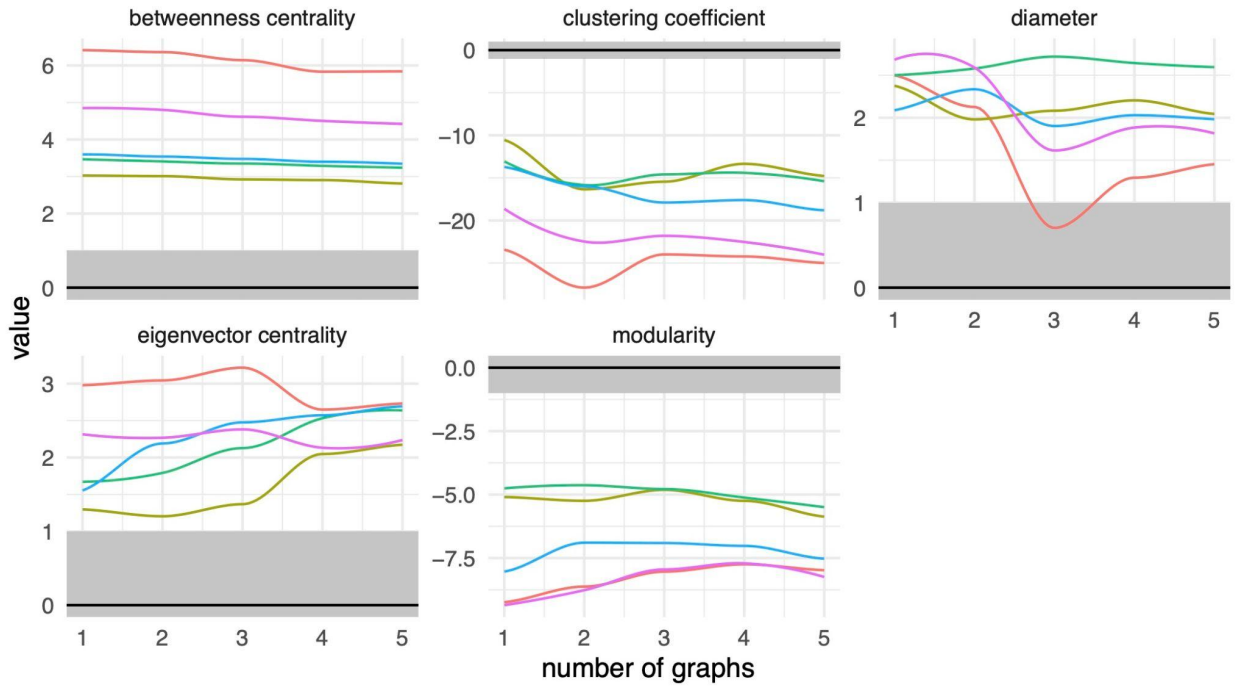

Figure 1. Comparison of empirical and null model values for the Etrurian settlement-road networks. Each line represents a set of time-averaged networks created based on a different original graph. The gray area in ±1 refers to ME values where the empirical values are not considered different from the null model values.

Chaco

For the Chaco dataset, the comparison of the time-averaged graphs to the null model shows that time-averaged values are all distinct from the null model for betweenness centrality, and mostly distinct for modularity, and eigenvector centrality (see Figure 2). For clustering coefficient, we see few (9.7% - 21.6%) networks falling within the model error range of the null model, indicating that some original and time-averaged networks do not produce values that are distinguishable from random networks. However, this is highly dependent on the original

network that is used for the basis of time-averaging. The network diameter values overlap with the null model values throughout all amounts of time-averaging. Specifically, the diameter is not distinct from the null model values for 56.8% to 77.5% of the cases. This trend is more frequent with no or low amounts of time-averaging where the diameter overlaps with the diameter of a random network for more than 70% of the cases and then slightly reduces to less than 60% for high amounts of time-averaging. Betweenness centrality appears to be relatively robust to time-averaging, with distinct signals being maintained at relatively the same value even as time-averaging increases. For the same reason, clustering coefficient, modularity and eigenvector centrality also seem to be relatively robust for some of the original networks.

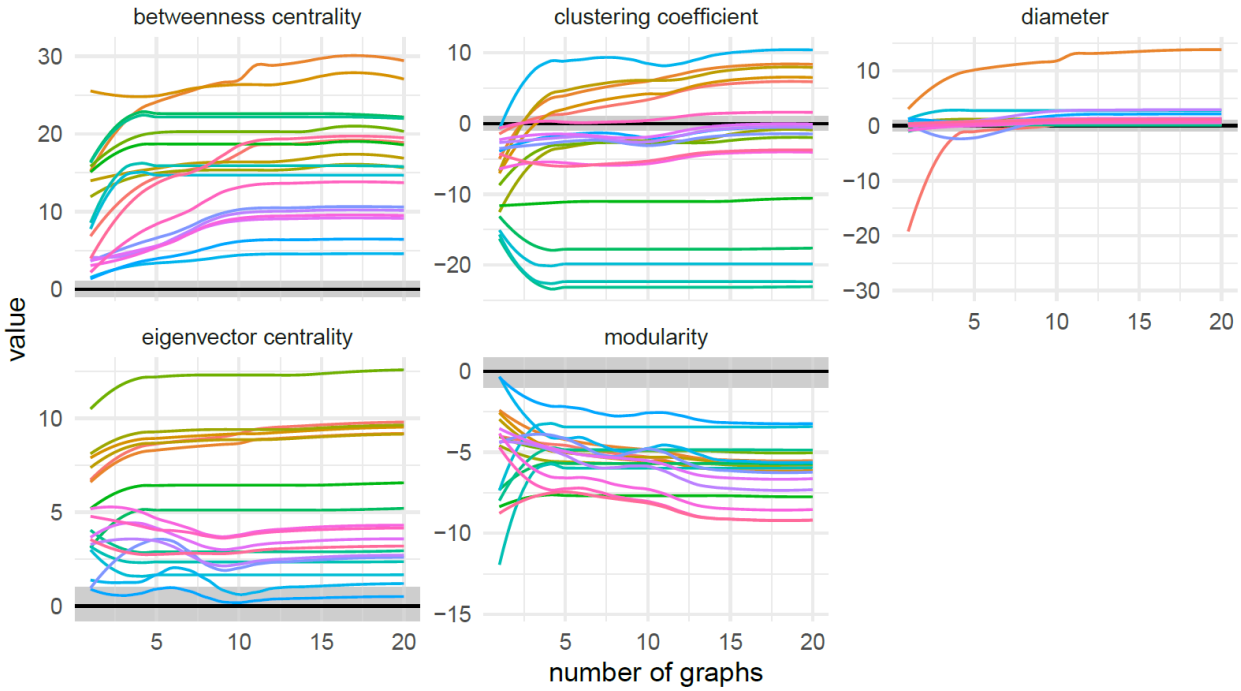

Figure 2. Comparison of empirical and null model values for the Chaco networks. Each line represents a set of time-averaged networks created based on a different original graph. The gray area in ±1 refers to ME values where the empirical values are not considered different from the null model values.

ICRATES

In the case of the ICRATES dataset, the effects of time-averaging are more complicated to assess (see Figure 3a-b). This is due to the nature of the dataset which consists of a series of pottery wares with a small number of wares strongly dominating the overall assemblage counts. This results in highly connected graphs with only a few clusters for each network slice.

Because the ICRATES networks are so strongly interconnected, betweenness centrality is very low. Yet, as can be observed from Figure 3b, the observed betweenness centrality is still largely distinct from randomized network behavior. Similarly, the diameter of largely connected

graphs will always be small, which is why there are only a few graphs from the ICRATES dataset that could be reasonably compared to a null model in terms of diameter.

Both in betweenness centrality and diameter, we obtain a model error value of “infinity”(thus skewing the respective graphs in Figure 3A) for 29.2% to 34% and for 45.7% to 95.6% of the networks, respectively (see Figure 3B). This occurs because the randomization process has no effect on the overall structure of fully-connected networks, returning values identical to the original network values, thus the nominator of ME (see Materials and Methods) will be zero and our resulting value will be infinite.

All the other metrics variably overlap with the null model of the original graph, suggesting that the effects of time-averaging are dependent on the initial network structure. Clustering coefficient values are distinct from random as only about 0.7% to 5.2% overlap with the null model values. Additionally, within each group of time-averaged graphs (represented by the differently colored lines in Figure 3), clustering coefficients seem to be relatively stable for most of the cases as time-averaging increases. However, for some sets of time-averaged graphs, time-averaging appears to increase the clustering coefficient value. A similar pattern is observed for eigenvector centrality, but. The non-random signals of eigenvector centrality are lost in more cases compared to the clustering coefficient. In 10.9% to 30.4% of the cases, the eigenvector centrality value overlaps with the null model values (see Figure 3b). Regarding modularity, the empirical values are not different from the null model for small to middle amounts of time-averaging for almost 50% of the cases, while for middle to high amounts of time-averaging, the empirical values are distinct from the null values and become more distinct as the amount of time-averaging increases. Specifically, when 25 slices are aggregated, 64.4% of the network values are distinct from the null model values and this portion steadily increases to 90.1% when 33 slices are aggregated, while more than 95% of the networks have a modularity value distinct from null when more than 40 slices are aggregated. Of all the network metrics applied here, the clustering coefficient appears most robust for ICRATES.

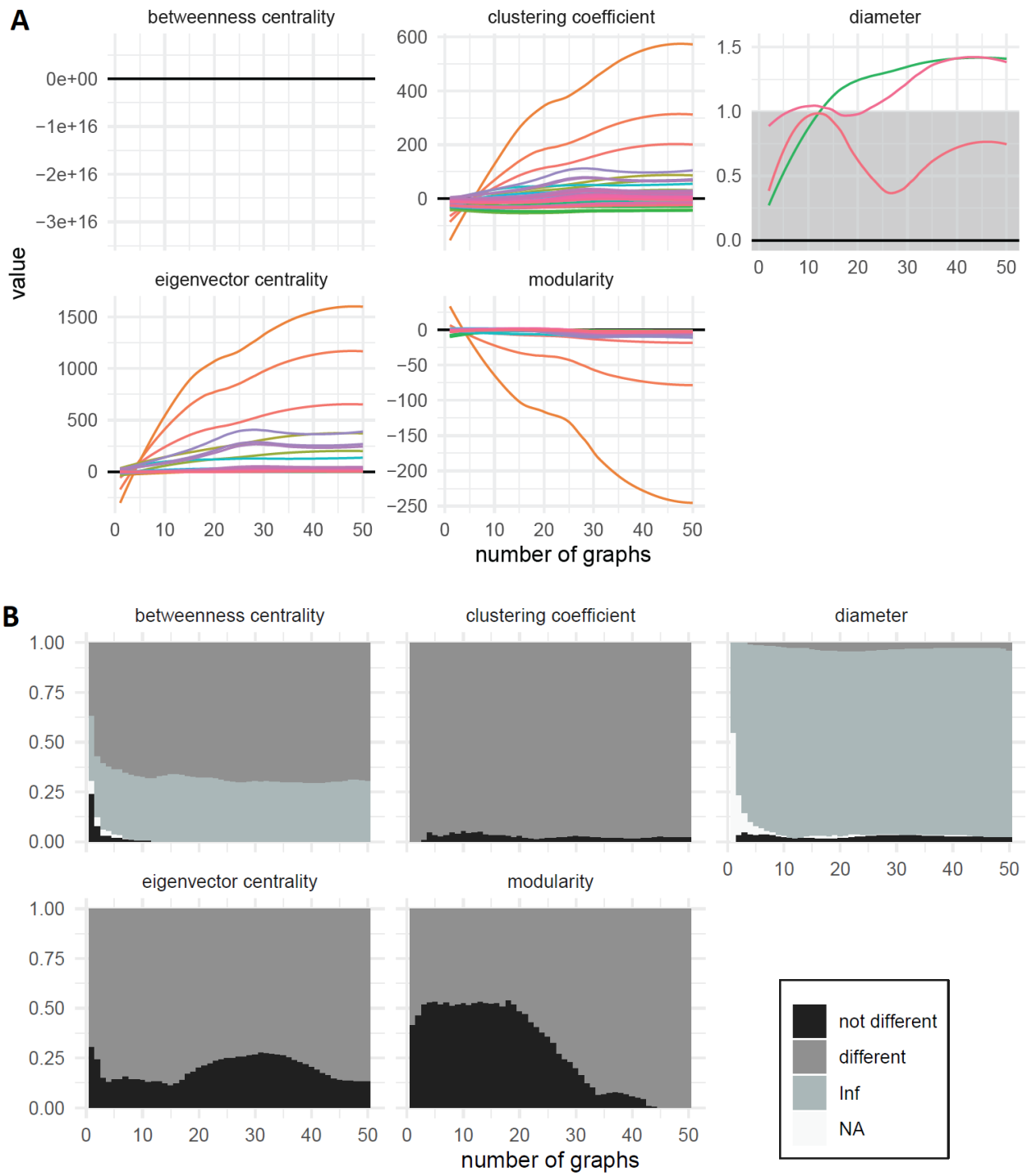

Figure 3 a-b. Comparison of empirical and null model values for the ICRATES networks. In figure a, each line represents a set of time-averaged networks created based on a different original graph. The gray area in ±1 refers to ME values where the empirical values are not considered different from the null model values. In figure b, we plot the percentage of values per metric that overlap with the null model value for increasing amount of time-averaging.

Original versus time-averaged networks

To further understand how time-averaging alters network topology, we compared time-averaged network metrics to the values of each original network. Through principal component analysis (PCA), we created an overview of the metrics for all time-averaged networks using a common coordinate system. As input data for PCA we only used the values of the time-averaged networks while the original network values are projected on the computed principal axes. We included the number of graphs as a variable in the analysis because it captures the temporal sequence and therefore helps to achieve independence between observations. The results are summarized using a PCA biplot. Finally, having mapped the multivariate data in the PCA biplot, we proceed with the comparison of time-averaged and original networks. We visually evaluate the correlation between metrics in the first two components and the relative position of the points referring to the original networks *versus* the time-averaged ones.

Etruria

For the Etrurian settlement-road dataset, the amount of time-averaging is positively correlated with betweenness centrality. The amount of time-averaging is uncorrelated with the clustering coefficient, and modularity in PC1 and PC2; therefore, these metrics are less prone to changes induced by aggregating network slices for this dataset. The fact that points representing networks with low time-averaging are spatially closer to the original networks signifies that higher time-averaging distorts the retrieved metrics more. This is because as we reach higher amounts of time-averaging we combine graphs that are more dissimilar to each other, as can be seen by the separation between networks a and b compared to networks d and e in the biplot. Network c stands in between as low time-averaging distorts the network according to the direction of the time-averaging. Lastly, it should be noted that the PCA map of the Etrurian dataset explains 73.7% of the total variance between time-averaged networks, which is relatively low compared to Chaco (84.6%) and ICRATES (89.6%); therefore, a relatively large part of the variation is not depicted in the biplot.

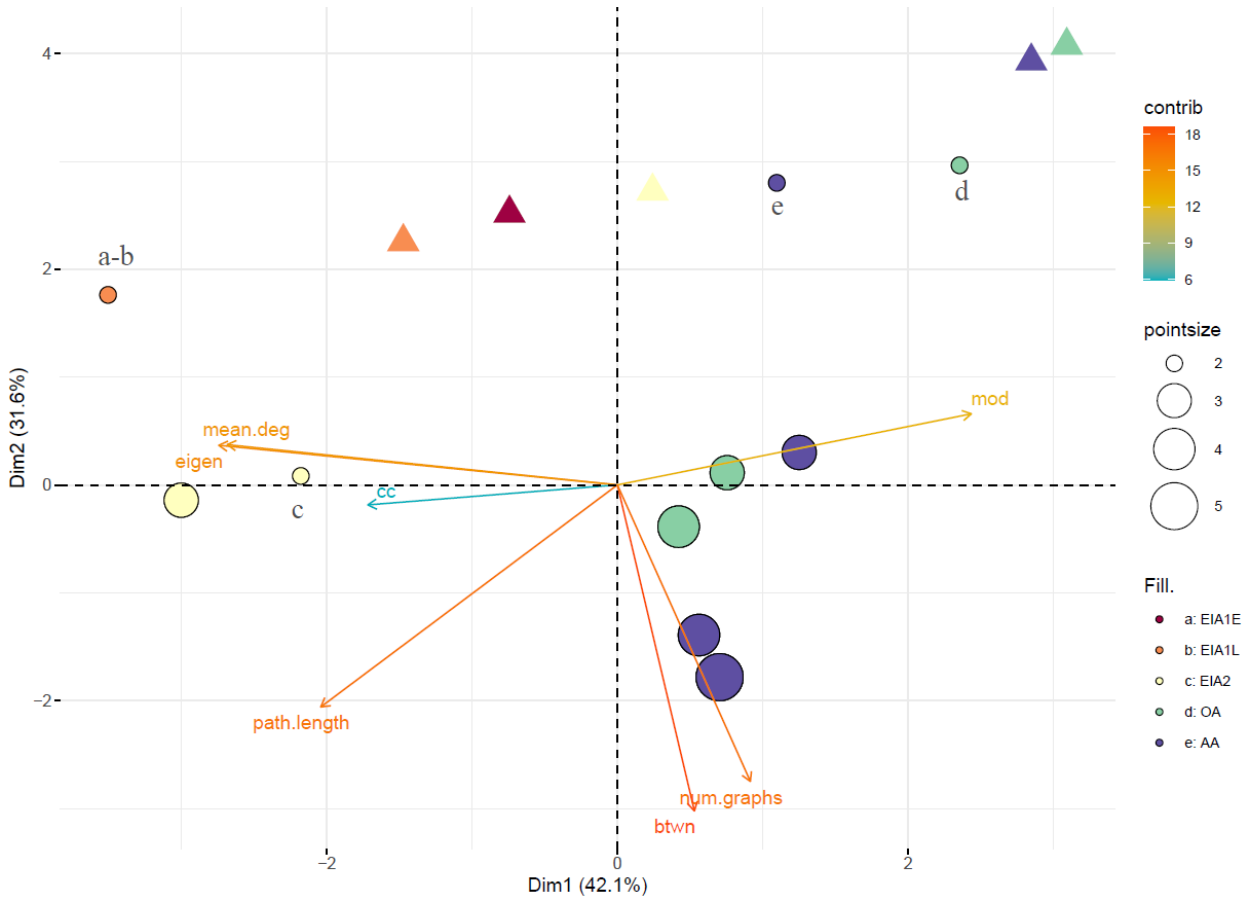

Figure 4. PCA with first two dimensions based on the network metric values of all time-averaged graphs for the Etrurian road-settlement networks. The circles show where on the PCA the time-averaged networks would be plotted. The lines represent the number of averaged graphs (num.graphs) and the network metrics included in the calculation of the PCA. These include mean degree (mean.deg), path length (path.length), clustering coefficient (cc), modularity (mod), eigenvector centrality (eigen) and betweenness centrality (btwn). The contribution of these variables to the PCA dimensions are given by the blue to red spectrum. The size of the points represents the amount of time-averaging and the color of the point represents the original graph used as a basis. The triangles represent the original networks.

Chaco

The amount of time-averaging in the Chaco networks is positively correlated with the values of modularity and negatively correlated with mean degree and clustering coefficient. Betweenness centrality and path length, on the other hand, are not strongly affected by the amount of time-averaging. Looking at the minimally-averaged networks of the Chaco dataset, similarly to the Etrurian data, we notice a pattern of structural difference between the networks retrieved from the early and the late original networks. Most networks (a-l) have negative scores on the first principal axis while the latest networks (m-t) have scores that are close to zero. In general, even low time-averaging distorts the structure of the original network, with the early

time period networks (a-b) and late period networks (q-t) being the most distorted, while c-p are clearly less distorted. Finally, original networks from the middle time periods (i-p) exhibit lower structural differences under scenarios of low time-averaging. Moreover, once initiated, higher amounts of time-averaging do not induce further distortions. Therefore, even highly time-averaged networks would not necessarily lead to critical misinterpretations when the goal is to study the period between 1000-1175 CE.

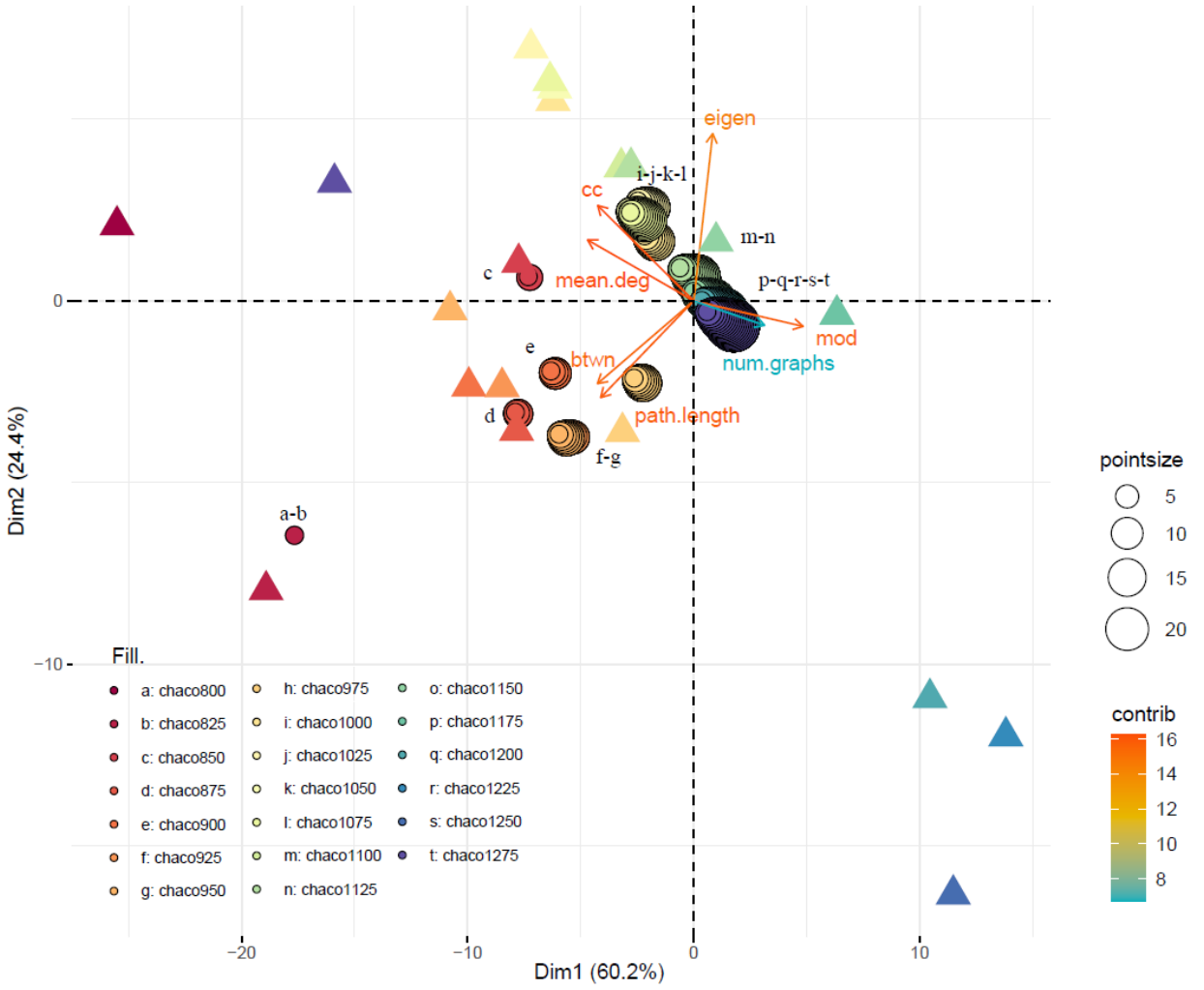

Figure 5. PCA with first two dimensions based on the network metric values of all time-averaged graphs for the Chaco networks. The circle points show where on the PCA the time-averaged networks would be plotted. The lines represent the number of averaged graphs (num.graphs) and the network metrics included in the calculation of the PCA. These include mean degree (mean.deg), path length (path.length), clustering coefficient (cc), modularity (mod), eigenvector centrality (eigen) and betweenness centrality (btwn). The contribution of these variables to the PCA dimensions are given by the blue to red spectrum. The size of the points represents the

amount of time-averaging and the color of the point represents the original graph which was used as a basis. The triangles represent the original networks.

ICRATES

In the ICRATES dataset, the amount of averaging is, as in Chaco data, positively correlated with the values of modularity and negatively correlated with mean degree, eigenvector centrality, clustering coefficient and path length. Interestingly, we also observe substantially higher degrees of overlap between the original networks and their averaged counterparts, even though this does not hold for all cases. In the latest networks (zv-zx), the structure is relatively well preserved only with a very small amount of time-averaging. At the same time, earlier networks (e-m) seem to maintain the structure of the non-time-averaged networks with a considerable amount of averaging but this is an artifact resulting from the fact that the first four networks have little to no connections. We also observe that the original network structures which are close in time, except for networks e-g and t-z, largely overlap in the biplot therefore are also close structurally. The original and the slightly averaged networks around 20 AC and 40 AC stand close to the biplot origin, which means these networks stand in the middle of the variations exhibited by the totality of the dataset and may reflect a transitional period. The highly averaged graphs are dissimilar to all original graphs. This means that studying the highly averaged graphs cannot provide meaningful interpretations regarding the network structure over a shorter time scale.

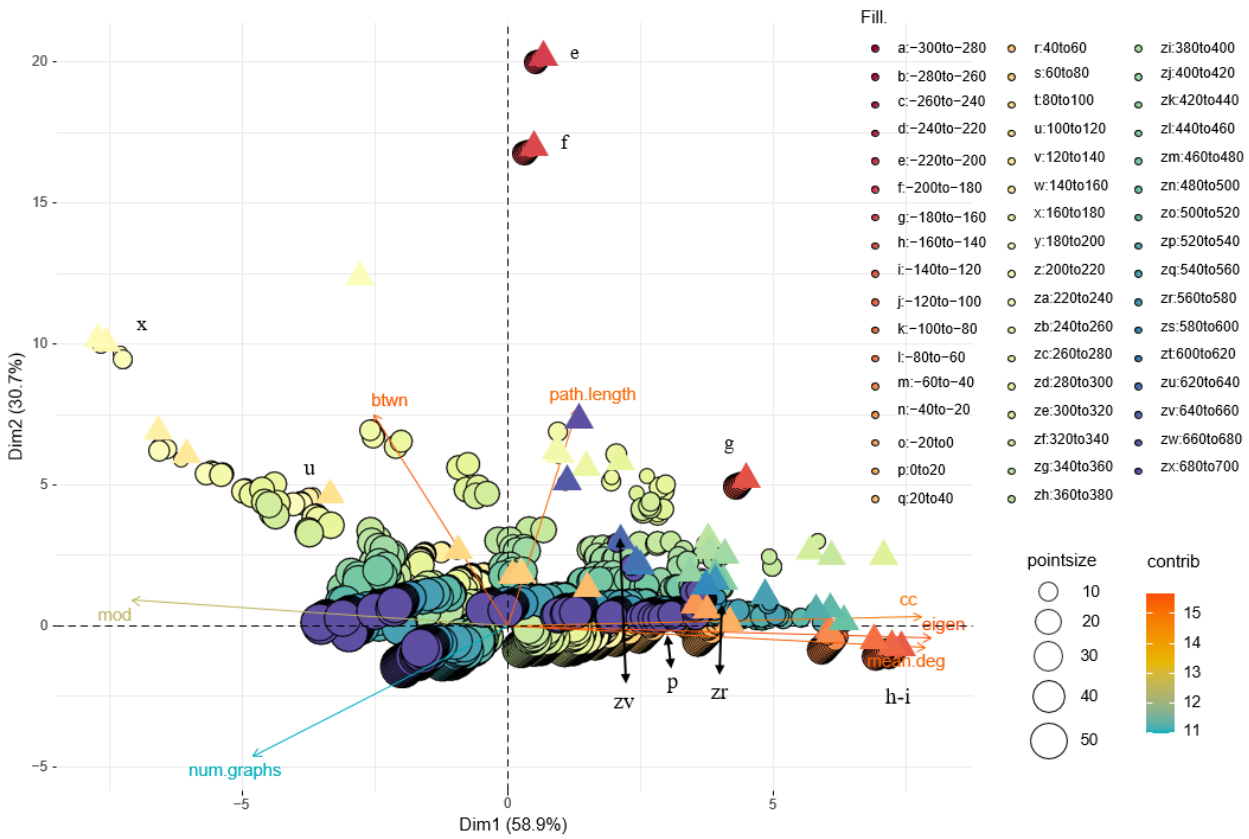

Figure 6. PCA with first two dimensions based on the network metric values of all time-averaged graphs for the ICRATES networks. The circle points show where on the PCA the time-averaged networks would be plotted. The lines represent the number of averaged graphs (num.graphs) and the network metrics included in the calculation of the PCA. These include mean degree (mean.deg), path length (path.length), clustering coefficient (cc), modularity (mod), eigenvector centrality (eigen) and betweenness centrality (btwn). The contribution of these variables to the PCA dimensions are given by the blue to red spectrum. The size of the points represents the amount of time-averaging and the color of the point represents the original graph which was used as a basis. The triangles represent the original networks.

Summary

Our analysis shows that the effects of time-averaging are contingent on the structures of the original graphs for the Etrurian and Chaco datasets, especially when the time-averaging is extensive. For ICRATES, a large amount of time-averaging is needed as a small amount of time-averaging does not necessarily alter the structure of the original graph. This appears to fall in line with the results discussed earlier that suggest finding general rules to help mitigate the effects of time-averaging might be a difficult endeavor. Still, it is clear that time-averaging generally affects the fidelity of information from networks to various degrees.

Time-averaging effects on node metrics

To fully assess the impact of time-averaging on archaeological networks, we supplemented our study of the ramifications on the level of the network as a whole, outlined in the previous parts, with an analysis of the implications on the level of the individual nodes and their position within the networks. Here, we specifically focus on centrality measures since those are most frequently used in archaeological network analyses.

Etruria

Our analyses of the Etrurian settlement-road network dataset show that all centrality metrics are affected by time-averaging. In general, we see a clear trend of decreasing similarity between node sets as time-averaging increases (see Figure 7). This can perhaps be explained due to the size of the graph. The settlement-road network dataset is the smallest of our three case studies, with at most 179 nodes per original network. This means that as we subject these networks to time-averaging we are increasing the number of nodes from which our highest- and lowest-value nodes are determined by a larger percentage compared to the other two datasets.

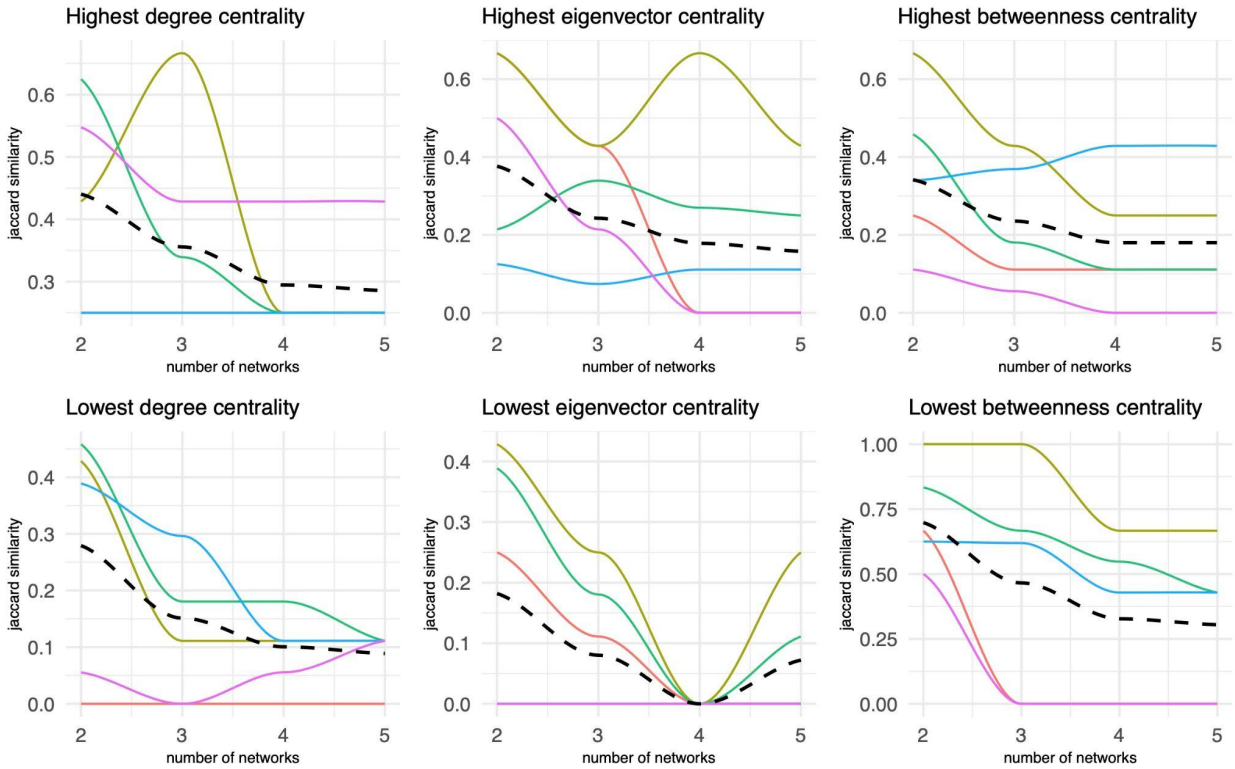

Figure 7. Similarity of the sets of nodes with extreme centrality values between the original network and each time-averaged network for the Euritrian networks dataset. The top three graphs display the results for the node sets with the highest centrality values; the bottom three graphs display the results for the node sets with the lowest centrality values. Each solid line

represents the node sets of time-averaged graphs, each based on a different original network. The dotted line represents the averaged Jaccard similarity values between the node sets across all the time-averaged graphs.

Chaco

For the Chaco dataset, we find a low similarity between low and high node sets for degree and eigenvector centrality when comparing time-averaged graphs to the original graphs (even though some variation for a couple of graphs can be observed) (see Figure 8). This means that the time-averaged graphs do not share any of the same extreme nodes with the original graph when using degree and eigenvector centrality. This suggests that eigenvector and degree centrality values are highly susceptible to the effects of time-averaging. In contrast, the similarity remains fairly high for the nodes with the lowest betweenness centrality across the different amounts of time-averaging. However, as we can see from Figure 8, there is a steep drop-off in similarity for all node sets when approximately 10 graphs are averaged. This pattern suggests that for the Chaco dataset, low amounts of time averaging do not significantly impact the interpretations that archaeologists might draw from node-based centrality measures.

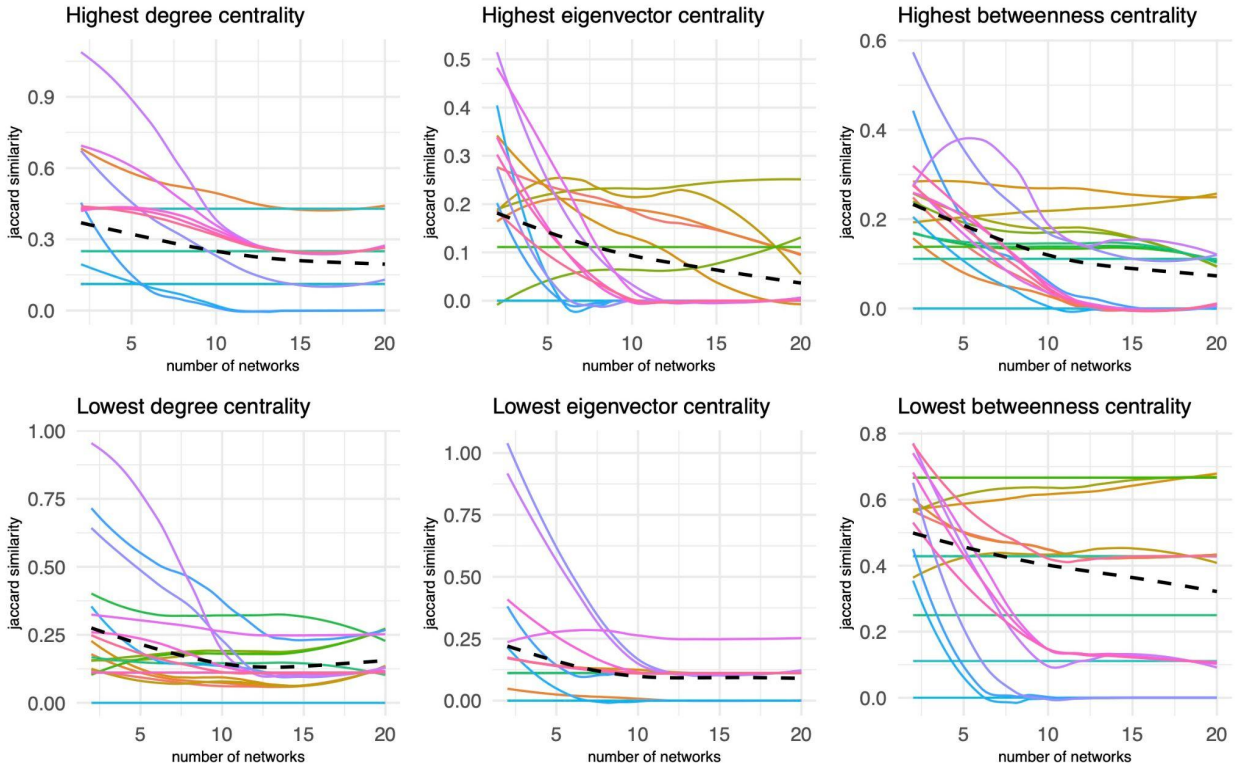

Figure 8. Similarity of the sets of nodes with extreme centrality values between the original network and each time-averaged network for the Chaco dataset. The top three graphs display the results for the node sets with the highest centrality values; the bottom three graphs display the results for the node sets with the lowest centrality values. Each solid line represents the node sets of time-averaged graphs, each based on a different original network. The dotted line

represents the averaged Jaccard similarity values between the node sets across all the time-averaged graphs.

ICRATES

In the ICRATES dataset, the lowest node values appear to be the least similar at extreme time-averaging (see Figure 9). The similarity value at the highest amounts of time-averaging is approximately 0.25 for high centrality nodes. The average trend is therefore one of decreasing similarity as time-averaging increases, with most significant changes occurring around the averaging of 10 or 20 network slices. We must note, however, that this is also highly dependent on which original graph we are basing the comparison on.

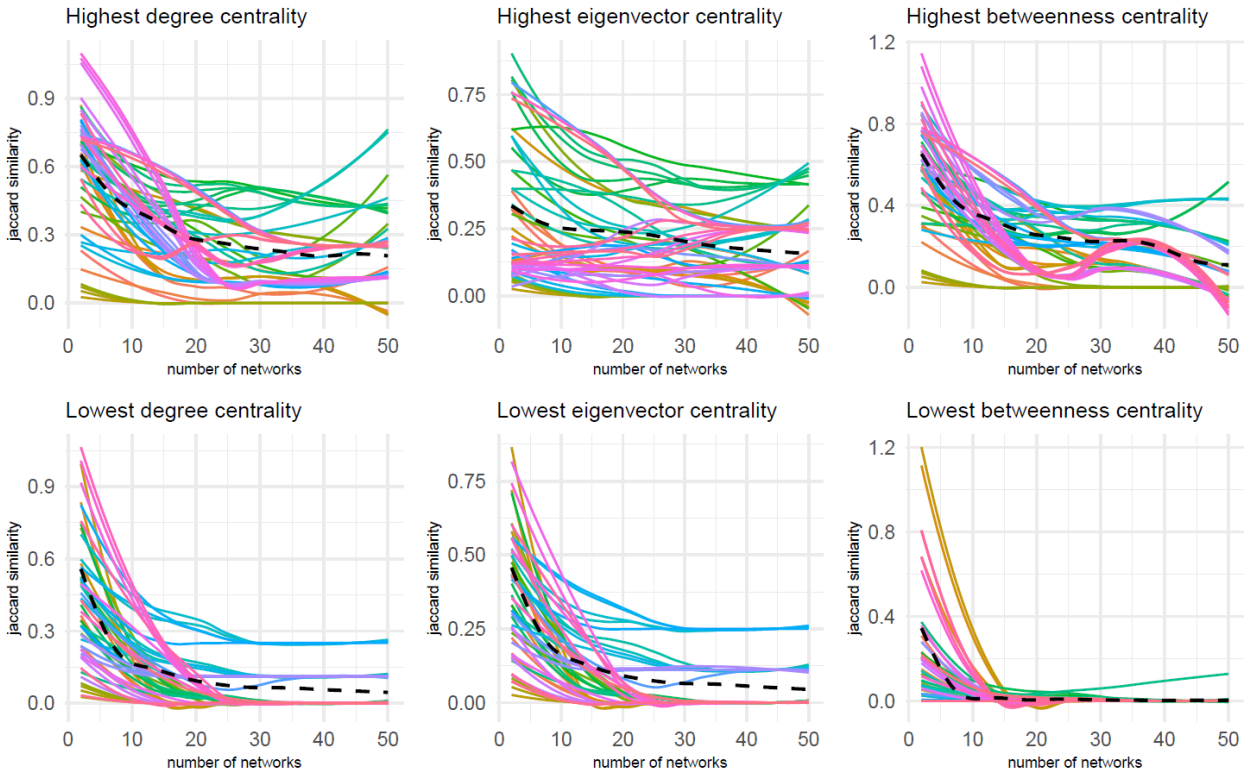

Figure 9. Similarity of the sets of nodes with extreme centrality values between the original network and each time-averaged network for the ICRATES dataset. The top three graphs display the results for the node sets with the highest centrality values; the bottom three graphs display the results for the node sets with the lowest centrality values. Each solid line represents the node sets of time-averaged graphs, each based on a different original network. The dotted line represents the averaged Jaccard similarity values between the node sets across all the time-averaged graphs.

Summary

Overall, the results of our node metric analysis clearly demonstrate that time-averaging significantly impacts relationships between nodes in a network. Because archaeologists typically use network-based analyses to understand interactions and relationships between sites, it is important to consider how the chronological resolution of these networks might play a role in obscuring the true relationships.

Sensitivity Analysis

For the sensitivity analysis, we focused on assessing two aspects of the original graphs and the time-averaged networks: the potential impact of missing nodes, on the one hand, and missing edges, on the other. We used five metrics (i.e., degree centrality, eigenvector centrality, betweenness, clustering coefficient, and modularity) to evaluate changes in networks as time-averaging increased. For degree, eigenvector and betweenness centrality we based our interpretation on the interquartile range (IQR) and the median of the 100 replicates of node and edge removal.

Potential Impact of Missing Nodes

As could be expected, the networks in the Etrurian settlement-road network display marked changes in each of the selected metrics as the sampling fraction increases (see Figure 10). In general, Spearman's Rho becomes more variable for all three centrality measures as the sampling fraction increases, while the median Rho value consistently decreases. However, we also find some distinct patterns for the individual metrics. For betweenness and degree centrality in particular, the IQR remains relatively stable with only minor deviations until 40%, followed by a sharper increase between 30% and 10%. For eigenvector centrality, however, we see a much more pronounced divergence in ranges from the very start of the subsampling process. Likewise, the median of the betweenness and degree centrality shows a much more gradual divergence across the subsampling process compared to eigenvector centrality, which is characterized by a much more pronounced drop. This suggests consistent responses to node removal across amounts time-averaging for the settlement-road network dataset. For clustering coefficient and modularity, we get similar results where the patterns seem to stay relatively stable across time-averaged networks, as levels across the sampling fractions are consistent between the different networks until 30%. Modularity is generally lower with the 100% sample and increases all the way to 10%. The clustering coefficient remains about the same until 60%, then decreases over the sampling fractions.

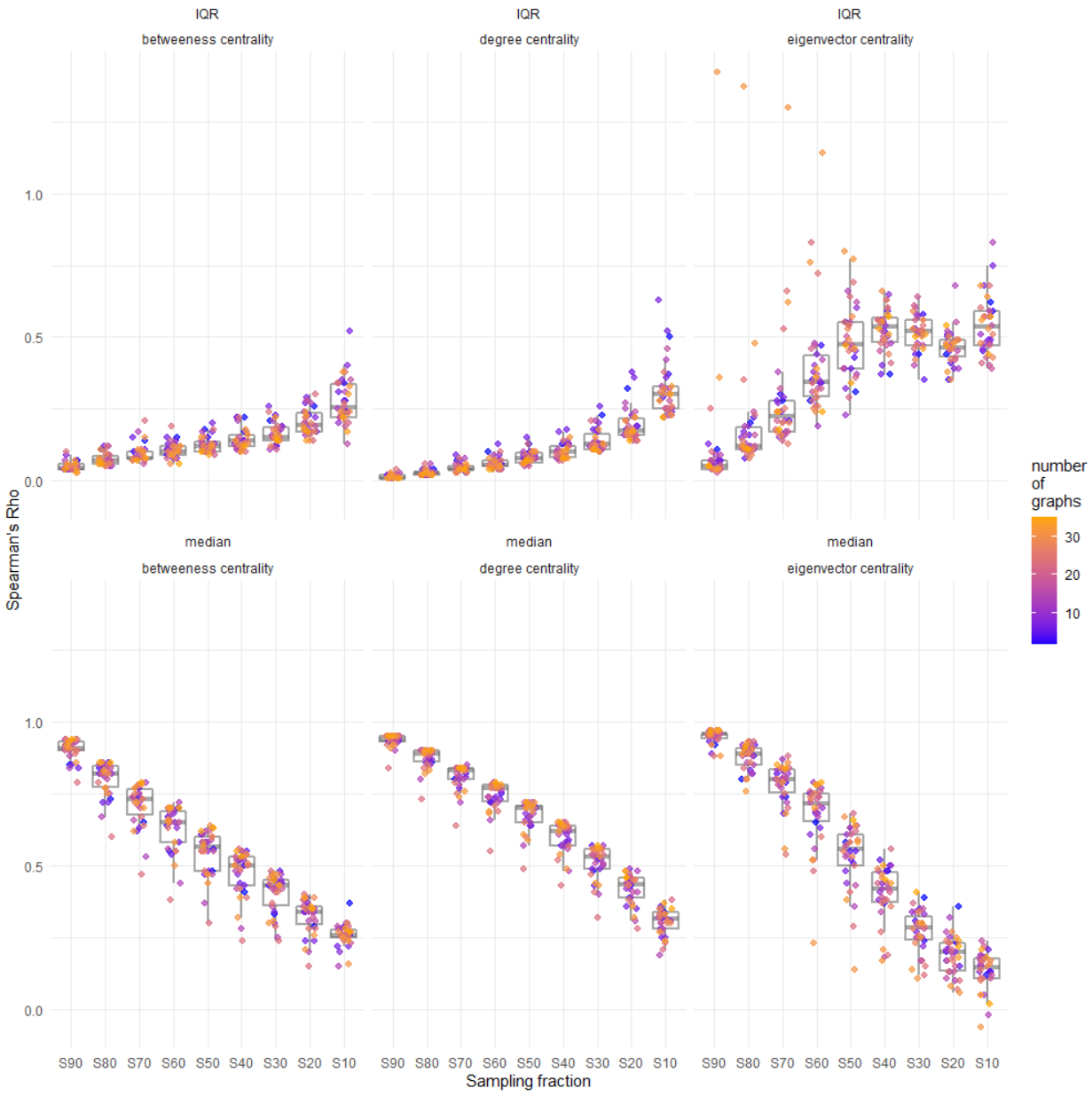

Figure 10. Summary graphs for sensitivity analysis on the nodes of Etrurian settlement-road dataset. Two statistics (median and IQR) are plotted for three metrics of centrality (betweenness, degree and eigenvector centrality) per sampling fraction. Points represent Spearman's Rho per statistic and metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

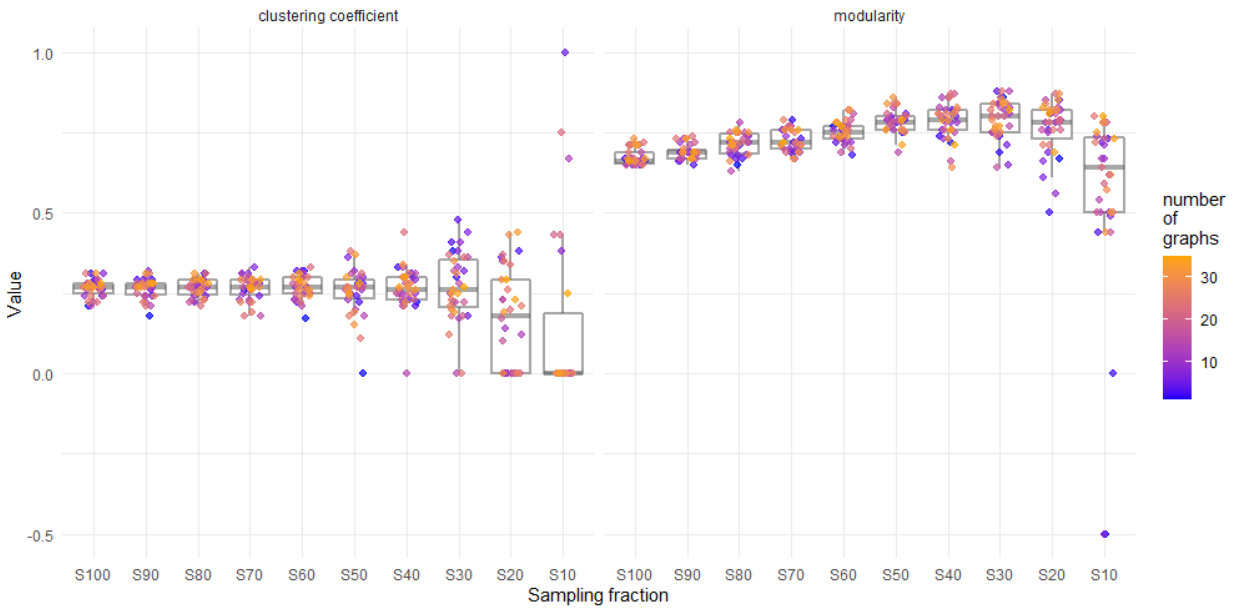

Figure 11. Clustering coefficient and modularity graphs for sensitivity analysis on the nodes of Etrurian settlement-road dataset. Points represent the metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

The results for the Chaco dataset show that degree centrality, which focuses on first-order connections, holds up well to node subsampling in the time-averaged networks.. For betweenness centrality, we see predominantly negative correlations between the original network metrics across all sampling fractions. Given that betweenness centrality is path-dependent, taking out a well-connected hub has a huge effect on the overall network as there is a high chance that many other nodes will see their shortest paths changed. For eigenvector centrality, we see very low correlations between the original network and the subsampled networks. This is perhaps unsurprising given the high clustering coefficients for this dataset. For all three centrality metrics, the median value remains fairly stable across all sampling fractions. Modularity remains fairly stable across the subsampling fractions, even though a gradual increase in the spread can be observed. The clustering coefficient likewise remains stable across all subsampling fractions, with an incremental increase in the spread of the values particularly from 30% onwards.

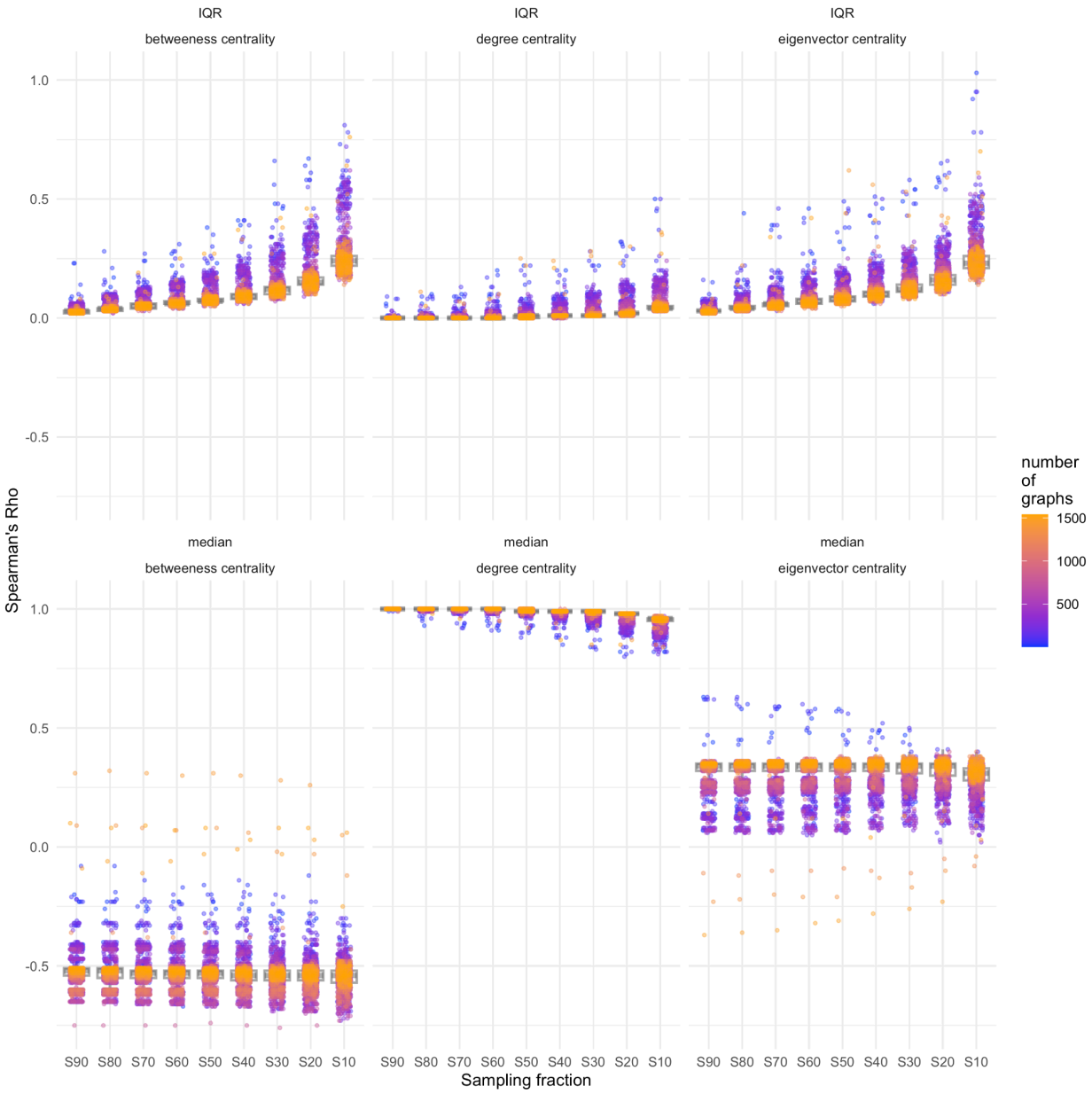

Figure 12. Summary graphs for sensitivity analysis on the nodes of Chaco dataset. Two statistics (median and IQR) are plotted for three metrics of centrality (betweenness, degree and eigenvector centrality) per sampling fraction. Points represent Spearman's Rho per statistic and metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

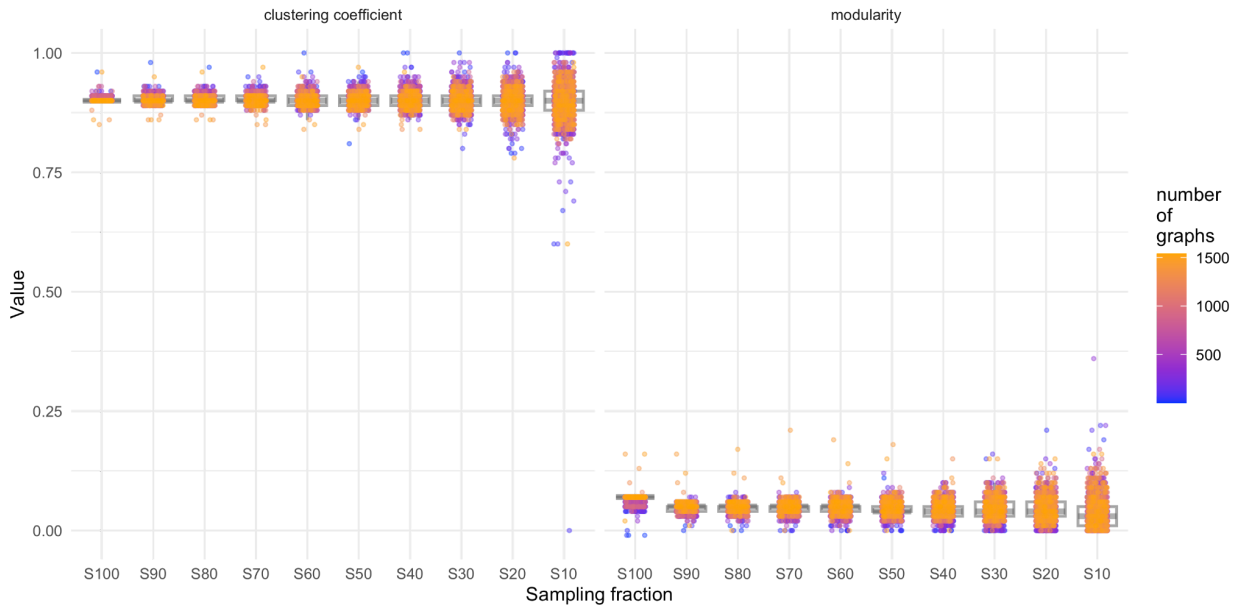

Figure 13. Clustering coefficient and modularity graphs for sensitivity analysis on the nodes of Chaco dataset. Points represent the metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

For ICRATES, median and IQR of the Spearman's Rho values for all three centrality measures remains stable across all subsampling fractions. The spread of the values increases at the onset of the subsampling process, but then remains fairly stable throughout the subsequent subsampling fractions. For betweenness and degree centrality, this initial jump is caused by a single group of networks that have little time-averaging and lower Rho values compared to the rest of the networks. For eigenvector centrality, the spread is notably higher compared to the other two, as it is caused by up to four deviating groups. If we zoom in from the macro-level to the individual networks, some periods yield more robust networks that hold up well to node subsampling, whereas others are less robust for degree, eigenvector, and betweenness centrality. In some cases where the original network is less robust to node subsampling, it actually becomes more robust with time-averaging. This effect is likely because the original networks often contain large fully connected components, which would be highly sensitive to any node removal. However, when the original graph is already robust to node subsampling, the robustness remains even when considering the time-averaged networks based on that original graph. In other cases, when starting out with a somewhat robust state (e.g., timeslice 400-420 CE), the network can become more robust with some time-averaging, but then experience decreasing robustness as time-averaging increases. For the clustering coefficient and modularity, the original graphs, generally, display a fairly steady level, albeit with a higher spread compared to the other two datasets. Interestingly, the clustering coefficient appears to be quite sensitive for time-averaging as the metric remains most stable for less time-averaged networks.

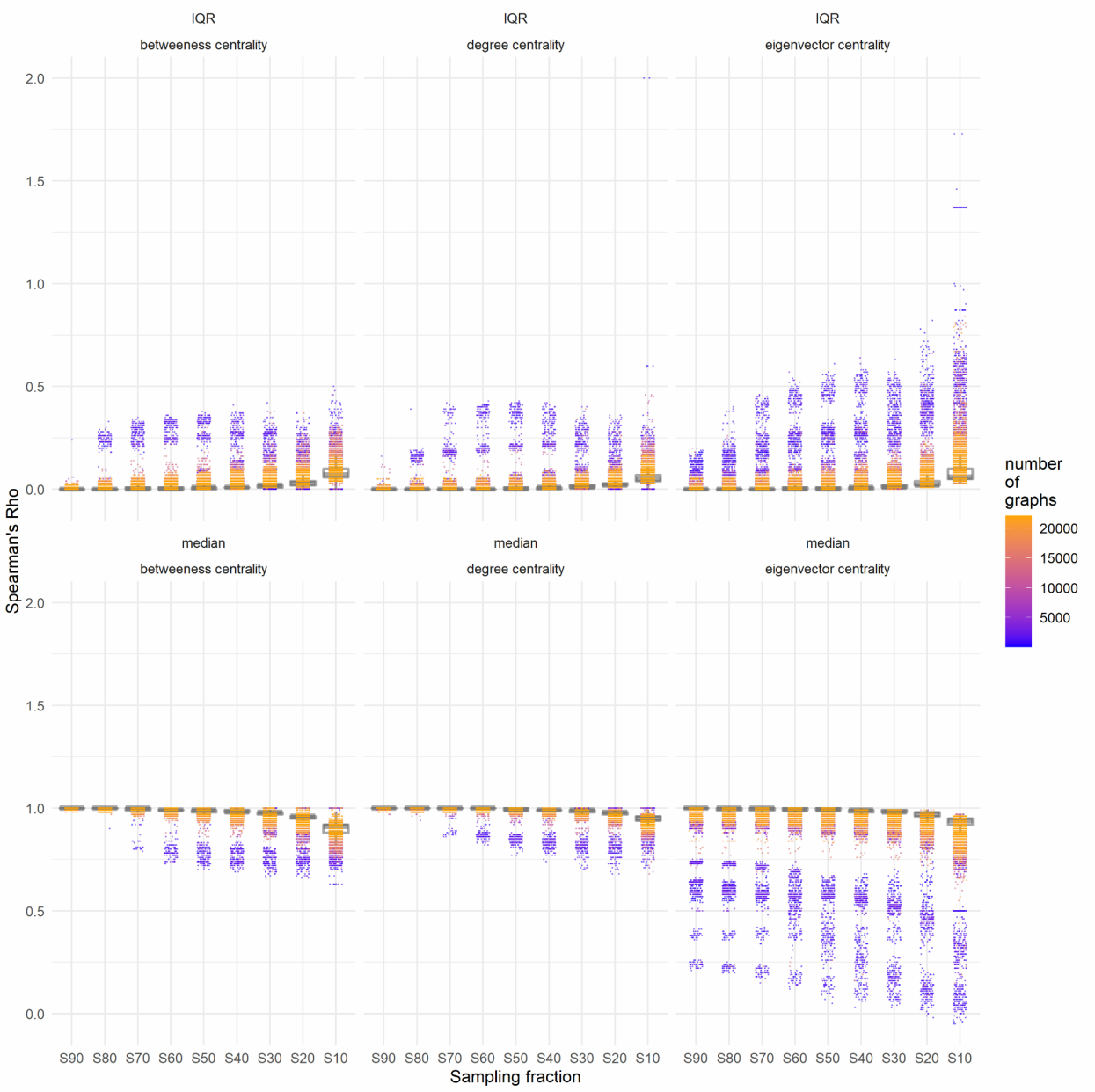

Figure 14. Summary graphs for sensitivity analysis on the nodes of ICRATES dataset. Two statistics (median and IQR) are plotted for three metrics of centrality (betweenness, degree and eigenvector centrality) per sampling fraction. Points represent Spearman's Rho per statistic and metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

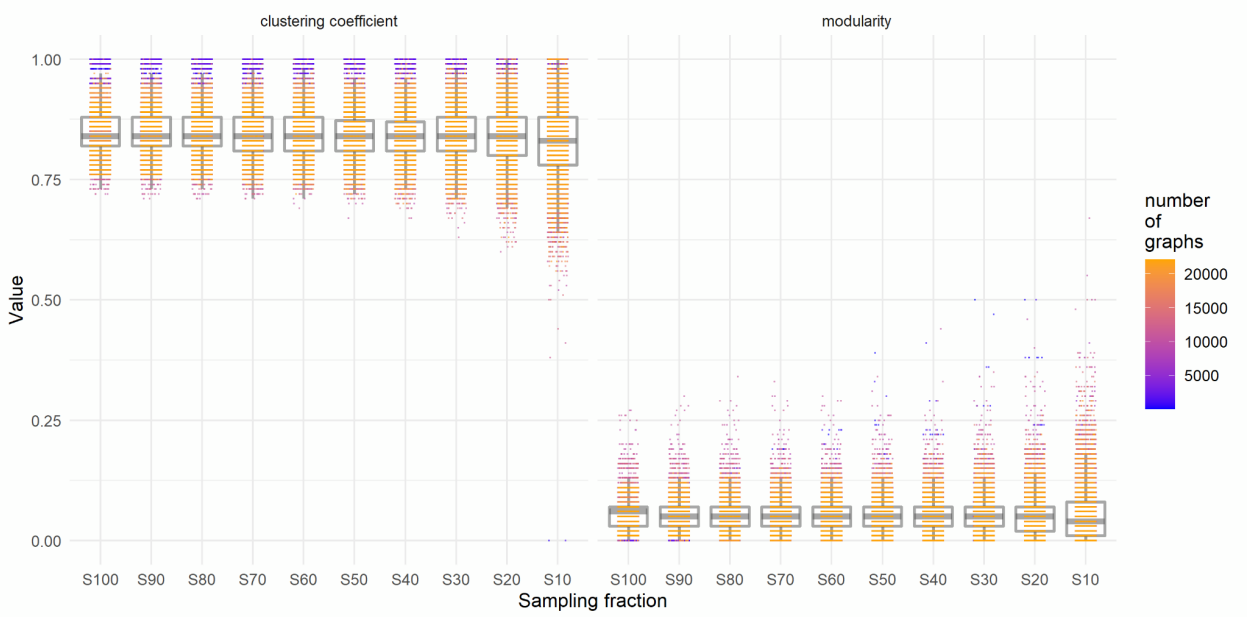

Figure 15. Clustering coefficient and modularity graphs for sensitivity analysis on the nodes of ICRATES dataset. Points represent the metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

Potential Impact of Missing Edges

The results of the Etrurian settlement-road network dataset show that edge removal can result in very similar patterns compared to the node removal results. Generally, there is a decrease in stability of network values across sampling fractions for each network for degree, eigenvector, and betweenness centrality, indicating that it is susceptible to edge subsampling. The median and IQR values of degree and betweenness centrality follow a linear trajectory, whereas for eigenvector centrality we see a lower rate of change until 60%, followed by a higher rate across the remaining subsampling fractions. However, as variability remains small, the effects of edge removal seem to remain relatively consistent for all five metrics when comparing the results from the original graphs to those from the time-averaged graphs. While the centrality measures do increase slightly with increased time-averaging, the change is overall fairly minimal and likely not very significant.

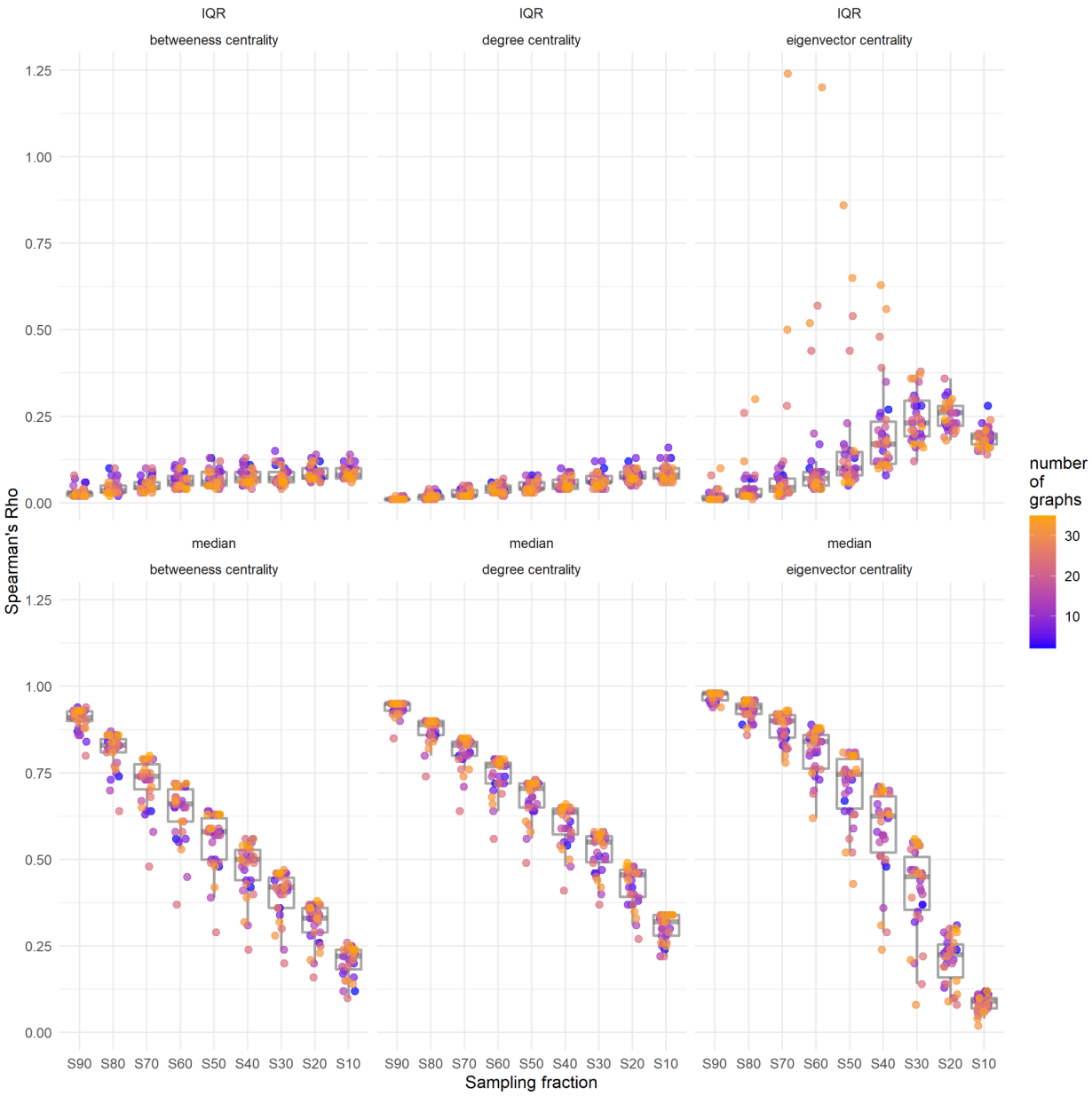

Figure 16. Summary graphs for sensitivity analysis on the edges of Etrurian settlement-road dataset. Two statistics (median and IQR) are plotted for three metrics of centrality (betweenness, degree and eigenvector centrality) per sampling fraction. Points represent Spearman's Rho per statistic and metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

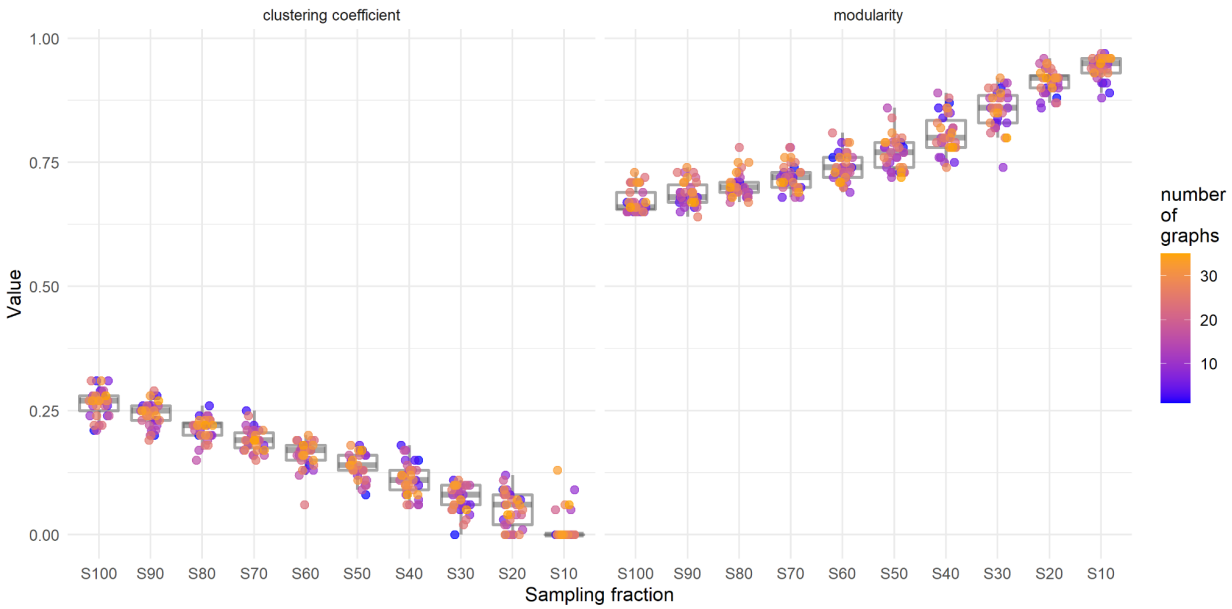

Figure 17. Clustering coefficient and modularity graphs for sensitivity analysis on the edges of Etrurian settlement-road dataset. Points represent the metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

For the Chaco dataset, the results of the edge removal procedure are quite different from those of the node removal results. In general, the Chaco networks are more robust to node subsampling compared to edge subsampling for degree centrality. The spread of betweenness centrality remains fairly stable, while also the median value is not much affected by the subsampling procedure. Until 50%, betweenness centrality remains more robust for edge subsampling compared to node subsampling, whereas from 20% it reverses and performs notably worse. Clustering coefficient performs worse in edge subsampling compared to node subsampling, whereas the opposite can be observed for eigenvector centrality. Modularity is the only metric affected in the same way for both node and edge subsampling. . For all three centrality metrics, the change in values is driven by a small group of networks with little time-averaging and lower Rho values. The clustering coefficient is high in the original graph and steadily decreases across the sampling fractions. The range of values also remains tight across sampling fractions and amount of time-averaging. Modularity stays fairly constant until around the 40% sampling fraction, after which a slight increase in values can be observed. This change seems to be driven by networks with low amounts of time-averaging.

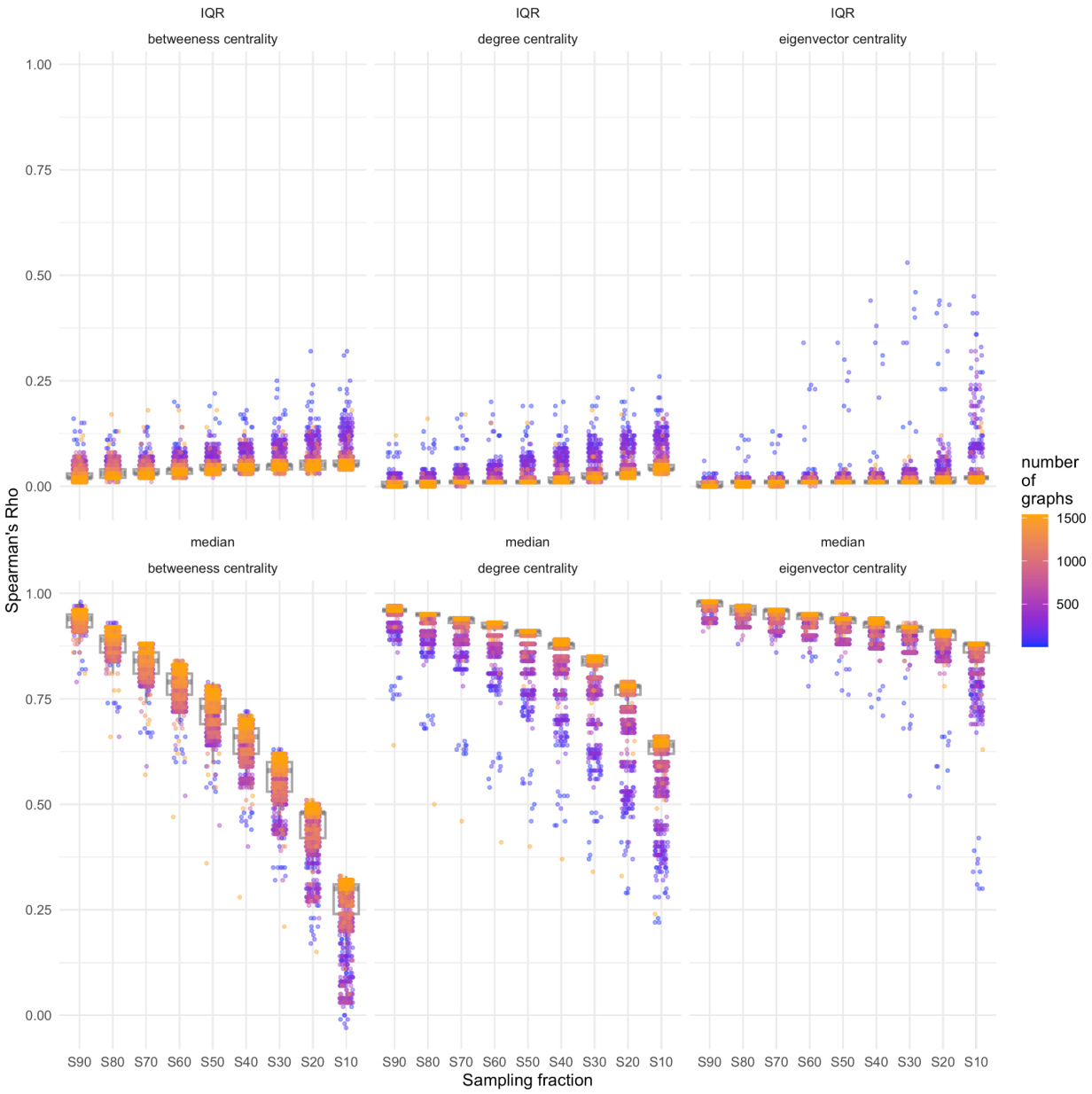

Figure 18. Summary graphs for sensitivity analysis on the edges of Chaco dataset. Two statistics (median and IQR) are plotted for three metrics of centrality (betweenness, degree and eigenvector centrality) per sampling fraction. Points represent Spearman's Rho per statistic and metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

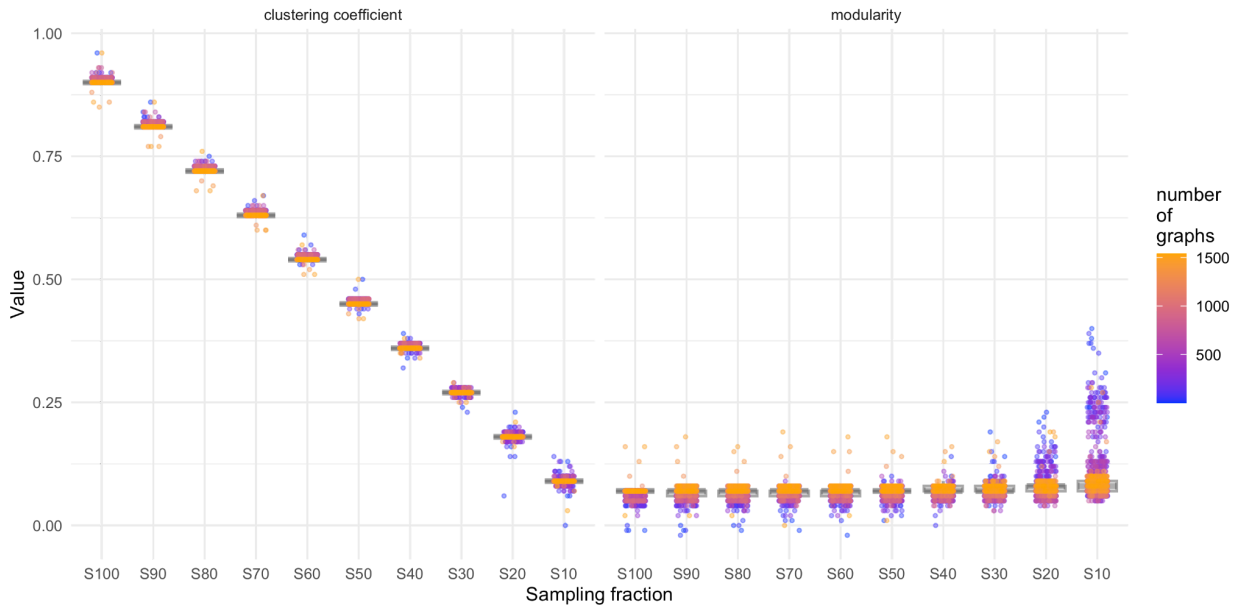

Figure 19. Clustering coefficient and modularity graphs for sensitivity analysis on the edges of Chaco dataset. Points represent the metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

As with the Chaco dataset, the outcomes of the edge removal procedure for ICRATES are quite different from the node removal results. In general, the networks are less robust to edge removal compared to node removal. The correlation of all three centrality metrics is notably affected even in high sampling fractions. This change is driven by three groups of networks with low time-averaging signatures, as well as a smaller component of the networks with medium to high time-averaging. Despite these diverging patterns, we find that the median value of all centrality metrics remains fairly stable until the sampling fraction 50% and slowly decreases afterwards. The patterns for modularity and clustering coefficient are relatively more stable (compared to the centrality metrics) across all time periods and degrees of time-averaging. Modularity stays very low until the 50% sampling fraction, then there is generally an increase until 10%. Interestingly, in this case we see a group of networks with mid-levels of time-averaging driving a mild divergence. Clustering coefficient gradually decreases throughout the time-averaging process. The low time-averaged networks largely show higher values, whereas the middle level networks decrease, followed again by an increase in the graphs with higher time-averaging. This overall pattern is maintained until 70%, after which it flattens out.

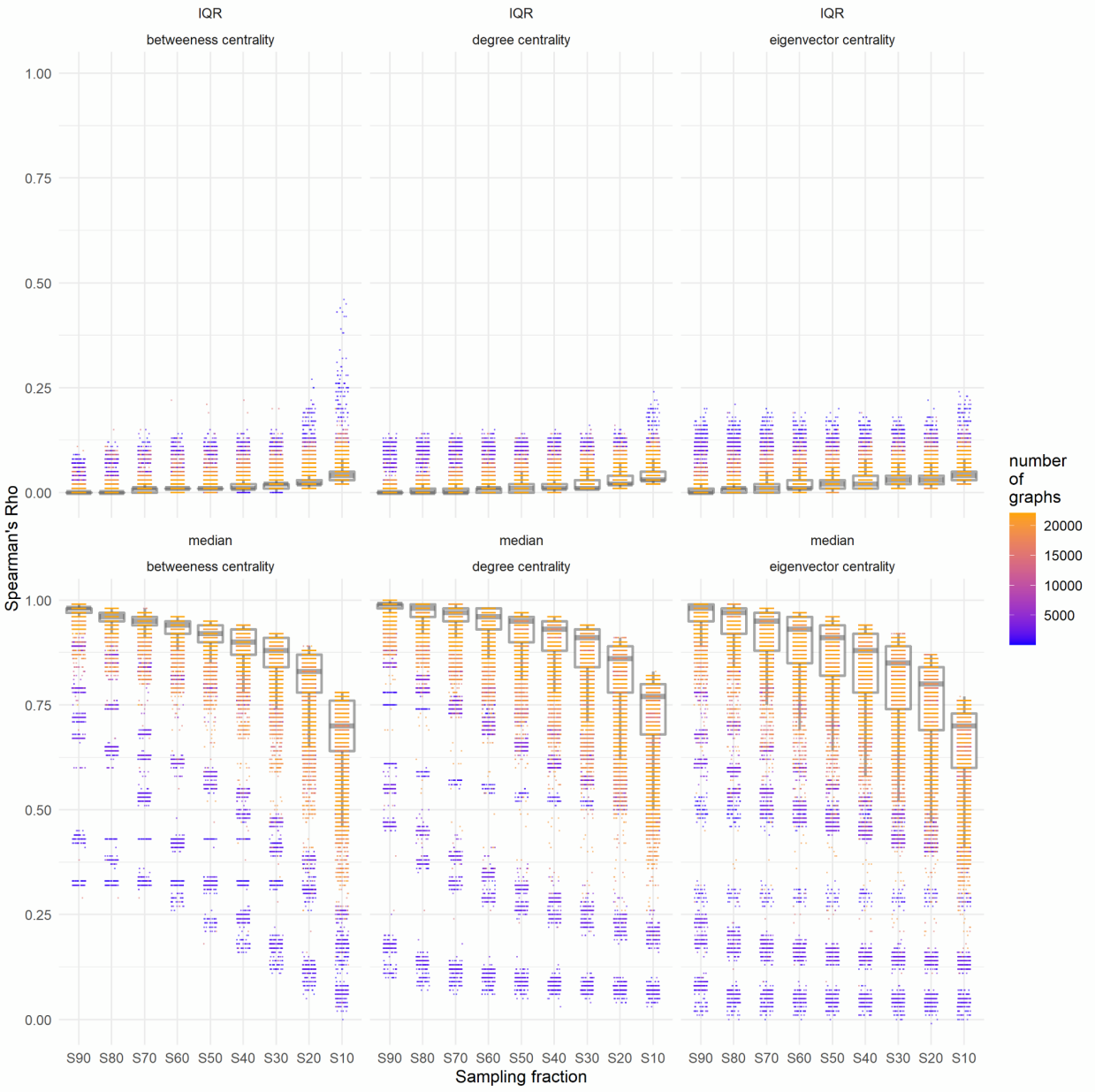

Figure 20. Summary graphs for sensitivity analysis on the edges of ICRATES dataset. Two statistics (median and IQR) are plotted for three metrics of centrality (betweenness, degree and eigenvector centrality) per sampling fraction. Points represent Spearman's Rho per statistic and metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

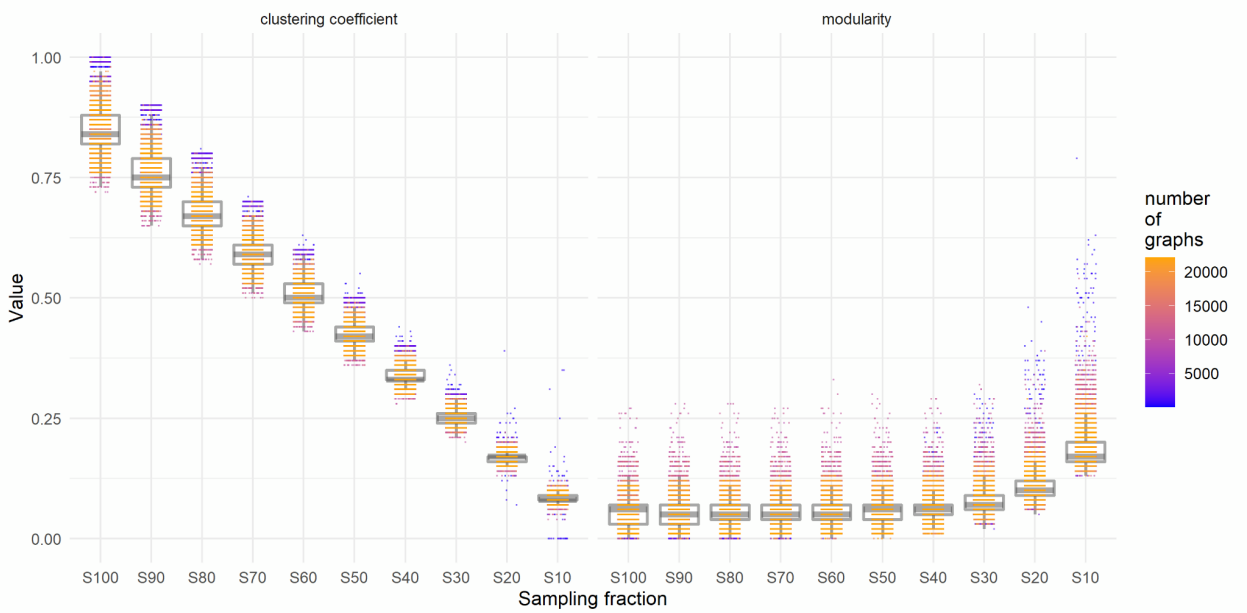

Figure 21. Clustering coefficient and modularity graphs for sensitivity analysis on the edges of ICRATES dataset. Points represent the metric value and are coloured based on the amount of time-averaging (number of graphs) imposed on the network they are acquired from.

Discussion

In this paper, we have explored the effects of time-averaging on three archaeological networks, both on the network and node level, as well as conducted sensitivity analysis to assess sensitivity to subsampling of nodes and edges in the original and time-averaged networks. Our analyses have shown that, while non-random signals of network topology can at times be maintained despite time-averaging, including information from multiple time slices in a single network can substantially impact the values of network science metrics, and therefore, also our interpretations of archaeological networks.

It is important to stress that with this analysis we do not evaluate the validity of our chronological data as such. On the contrary, the analysis makes no a priori assumptions regarding the accuracy of our data - both in scope and resolution - to allow independent assessment of the impact of time-averaging on each of the datasets. Not only would this be a monumental task in and of itself, it would detract from the main points this paper makes. This paper should therefore not be taken as proof or rebuttal of their chronological validity, but rather as a point of caution against the implications of using archaeological chronologies at face value, especially when using archaeological data as input for network analysis. As a related point, this means that the available chronological resolution (i.e. bin sizes of the network slices) plays an important role regarding the interpretation of the network at large. Take for example the noted discrepancy in resolution between the uneven time slices in the Etruria networks and the evenly distributed time slices in the Chaco and ICRATES datasets. The former is inherently more

coarse-grained and therefore already more highly, as well as more variably, time-averaged. This is a notable point of interest; however, the ramifications of using different bin sizes for the results of network analysis falls beyond the scope of this paper and will be discussed in future work.

The results of our analysis show that time-averaged networks are typically larger in size, with a larger diameter. All of the other metrics show that the impact of time-averaging on the overall network structures can vary greatly and in different ways (see Table 1 for an overview of results from network-level analyses). Interestingly, when we compare the two pottery networks based on the one hand on co-presence (ICRATES), and on the other on similarity (Chaco), we find that the correlations with amount of time-averaging have the same directions for all metrics (except eigenvector centrality), with differing magnitudes depending on the metric. The Etrurian road network patterns differ compared to the pottery networks in nearly all metrics, except for eigenvector centrality. This centrality measure is negatively correlated with the amount of time-averaging for the ICRATES dataset as well.

Table 1. Summary of the network-level analysis results. The values in column 'Distinct from null' summarize the results from the comparison of the networks to randomized graphs. The values in column 'Correlation with amount of time-averaging' summarize the results from the time-averaging procedure. The interpretation of correlation is based on Spearman's rho values; specifically, values [0, 0.3] are negligible, (0.3, 0.5] are weakly correlated, (0.5, 0.7] are moderately correlated, and (0.7, 1] are strongly correlated.

Dataset	Metric	Distinct from null	Correlation with amount of time-averaging
Etrurian	Betweenness centrality	Yes	Strongly positive
	Clustering coefficient	Yes	Negligible
	Diameter	Yes, most graphs	------
	Eigenvector centrality	Yes	Weakly negative
	Modularity	Yes	Negligible
	Mean degree	------	Negligible
	Path length	------	Negligible
Chaco	Betweenness centrality	Yes	Weakly negative
	Clustering coefficient	Yes, most graphs	Weakly negative
	Diameter	No, few graphs	------
	Eigenvector centrality	Yes, most graphs	Negligible
	Modularity	Yes, most graphs	Moderately positive
	Mean degree	------	Moderately negative
	Path length	------	Weakly negative
ICRATES	Betweenness centrality	Some graphs	Negligible
	Clustering coefficient	Yes, most graphs	Moderately negative

	Diameter	No, few graphs	------
	Eigenvector centrality	Some graphs	Moderately negative
	Modularity	Some graphs	Moderately positive
	Mean degree	------	Weakly negative
	Path length	------	Strongly negative

Archaeologists use networks for a variety of different purposes ranging from spatial to conceptual and computational networks (Knappett 2021; Östborn and Gerding 2014). Time-averaging can add more nodes and edges to a network, which can have varying impacts. For example, this could increase the chance of a central hub being subjected to “reflection” when calculating eigenvector centrality, as during the calculation the centrality is passed to neighbors and then inevitably gets passed to the central node again (Martin et al. 2014). Our own results demonstrate low similarity between node sets identified as having high centrality values, especially with large amounts of time-averaging. Furthermore, as the summarized results in Table 1 show, the effects of time-averaging can vary widely for different types of networks and datasets. For example with Chaco, the mean degree of a network is moderately negatively correlated with time-averaging, which means that we are likely to underestimate the average number of edges per node in the graph with increased time-averaging. Table 1 also shows that none of the measures have the exact same correlation result between all 3 datasets. Depending on the network and the nature of the time-averaging, we could end up over- or under-estimating different results, such as the importance of a pilgrimage site (Mills 2017).

It becomes clear from our survey of these three different archaeological networks that the effects of time-averaging are highly dependent on initial network structures. This makes it difficult to establish general rules for how to interpret time-averaged networks in archaeology. It also emphasizes the importance of tailored approaches to assess and address time-averaging in individual case studies.

Still, we can provide the following general recommendations for evaluating the effects of time-averaging in archaeological networks. First, it is important to assess the nature and structure of the dataset used for network analysis. A useful framework is offered by Perreault’s (2019) discussion of the quality of archaeological datasets in terms of scope, resolution, sampling interval, and dimensionality. This evaluation allows archaeologists to assess whether the questions they ask are underdetermined by the archaeological record or not. That is, whether the quality of our observations is sufficient to distinguish between multiple potential processes generating the archaeological record. It is essential in this regard to assess the potential impact of time-averaging on each of these three parameters.

Second, it is advisable to explore the effects of time-averaging on your datasets for both network- and node-level metrics. Archaeological interpretations often revolve around understanding of the whole of the network as well as the role of its most important nodes; for example, network analysis is used to assess the role of regional capitals in settlement networks. When doing analyses such as these, we must ensure that the chronological signal of the data is robust prior to employing network-based approaches.

Third, comparisons with null random networks are highly informative to assess which metrics can provide meaningful insights into the structure and properties of your network. Given that the effects of time-averaging are highly contingent on the structures of the original graphs, the degree of robustness for each metric can differ from case to case. Understanding which metrics allow greater fidelity in the results is essential to build stronger interpretations and draw relevant conclusions from the dataset.

Fourth, sensitivity analysis can likewise provide useful insights into the robustness of an archaeological network and the impact of potential missing nodes and links due to partial preservation and other biases. In fact, due to the fragmentary nature of the archaeological record, sensitivity analysis should be part and parcel of all analytical studies in our discipline.

Finally, the results of the Etrurian network dataset and the Chaco dataset analyses suggest that certain aspects of network topology can remain relatively unaffected by time-averaging. For example, the apparent robustness of average betweenness centrality of the network for both datasets may suggest that these networks have a few central nodes that are consistent across all time intervals. This could indicate that networks with large hubs better resist the effects of time-averaging.

Conclusions

Time-averaging is an important process affecting the very core of archaeological data: its reliability for assessing temporal patterns. Yet, its effects often go unrecognized or are underestimated. Studies that do reflect on the effects of time-averaging focus on particular processes such as “taphonomic bias” favoring recent material traces over older ones (Surovell and Brantingham 2007; Surovell et al. 2009), or the spatial scale of cultural similarity and differentiation in archaeological assemblages (Miller-Atkins and Premo 2018). Only a few studies assess the general effects of time-averaging in archaeology, for example through the rate of change observed in the archaeological record (Perreault 2018).

None of the studies consider the effect of time-averaging in the usage of archaeological material for the creation and analysis of archaeological networks. Our analysis has shown that time-averaging strongly reduces the fidelity of network interpretations in significant, but highly variable ways, with effects being highly contingent on the structure of the original networks. This makes it difficult to develop blanket solutions and must be considered as a warning that each study must explore the potential effects of time-averaging in its own datasets. This paper offers a methodological pipeline, consisting of time slice creation, time-averaging and sensitivity analysis, which can be used by other researchers in assessing the effects of time-averaging in their own datasets and analyses. The accompanying code is openly available and provides everyone the right tools for this analysis. Through this paper, we aim to advance the ways that archaeologists and other historical scientists deal with the effects of time-averaging in their studies, especially when conducting network analysis, thus raising the bar and amplifying our fidelity in temporal data for archaeological and historical studies.

Acknowledgements

We are grateful to the organizers of Santa Fe Institute's Complex Systems Summer School (class of 2019), where we first discussed the ideas for this paper. We would also like to thank the many participants that we had discussions with and all the lecturers for CSSS 2019. More specifically, we thank Jack Shawn, Kate Wootton, and Anshuman Swain for their collaboration and discussions on network metrics and time-averaging. Thanks to Archaeology Southwest and Matt Peeples for providing the Southwest Social Network (SWSN) Database 1.0 data used in this article. We thank Matt Peeples and one anonymous reviewer for their thoughtful comments and helpful suggestions. This work was also supported in part through the NYU IT High Performance Computing resources and services.

Statements and Declarations

Author credit statement

DD, EC and AG-B contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. The first draft of the manuscript was written by DD, EC and AG-B, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Funding

DD was supported by the Academic Foundation Leuven to attend the Complex Systems Summer School in 2019 and by C1 funding [C14/17/025] from KU Leuven. EC and AG-B received no funding for conducting this study. DK received funding from the Research Foundation Flanders [G088319N].

Data availability

All code and network data used in the analysis for the current study are available in a GitHub repository (https://github.com/cocoemily/time-averaged-networks) and as part of an Open Science Framework project (https://osf.io/5ca2u/). Links to the original datasets of ICRATES and the Etrurian settlement-road network have been provided in the text. The Chaco data can be

accessed through the Southwest Social Network (SWSN) Database:https://cyberSW.org. In the repository we include only the aggregated network slices.

Supplementary materials

SI files can be found in the OSF repository (https://osf.io/5ca2u/).

References

Allaire, J. J., Xie [aut, Y., cre, McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., Iannone, R., Dunning, A., filter), A. Y. (Number sections L., Schloerke, B., Sievert, C., Dervieux, C., Ryan, D., Aust, F., Allen, J., Seo, J., Barrett, M., Hyndman, R., Lesur, R., Storey, R., Arslan, R., Oller, S., RStudio, PBC, inst/rmd/h/jqueryui-AUTHORS.txt), jQuery U. contributors (jQuery U. library; authors listed in, library), M. O. (Bootstrap, library), J. T. (Bootstrap, library), B. contributors (Bootstrap, Twitter, library), I. (Bootstrap, library), A. F. (html5shiv, library), S. J. (Respond js, library), I. S. (highlight js, library), G. F. (tocify, templates), J. M. (Pandoc, Google, library), I. (ioslides, library), D. R. (slidy, library), W. (slidy, Gandy (Font-Awesome), D., Sperry (Ionicons), B., Drifty (Ionicons), StickyTabs), A. L. (jQuery, filter), B. P. J. (pagebreak L. and filter), A. K. (pagebreak L. (2021). rmarkdown: Dynamic Documents for R.

Arnold, J. B. (2019). ggthemes: Extra Themes, Scales and Geoms for “ggplot2.”

Auguie, B. and Antonov, A. (2017). gridExtra: Miscellaneous Functions for “Grid” Graphics.

Bes, P. (2015). *Once upon a Time in the East : The Chronological and Geographical Distribution of Terra Sigillata and Red Slip Ware in the Roman East*, Archaeopress Publishing Ltd.

Bes, P., Willet, R., Poblome, J. and Brughmans, T. (2019). Inventory of Crafts and Trade in the Roman East (ICRATES): database of tableware.

Brughmans, T. (2013). Thinking Through Networks: A Review of Formal Network Methods in Archaeology. *J. Archaeol. Method Theory* **20**: 623–662.

Brughmans, T. and Peeples, M.A. (2023). *Network Science in Archaeology. Cambridge Manuals in Archaeology*. Cambridge University Press, Cambridge, UK.

Buchanan, B., Hamilton, M. J., Kilby, J. D. and Gingerich, J. A. M. (2016). Lithic networks reveal early regionalization in late Pleistocene North America. *J. Archaeol. Sci.* **65**: 114–121.

Collar, A., Coward, F., Brughmans, T. and Mills, B. J. (2015). Networks in Archaeology: Phenomena, Abstraction, Representation. *J. Archaeol. Method Theory* **22**: 1–32.

Contreras, D. A. and Meadows, J. (2014). Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. *J. Archaeol. Sci.* **52**: 591–608.

Costenbader, E. and Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Soc. Netw.* **25**: 283–307.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal* **Complex Systems**: 1695.

Gjesfjeld, E. (2015). Network Analysis of Archaeological Data from Hunter-Gatherers: Methodological Problems and Potential Solutions. *J. Archaeol. Method Theory* **22**: 182–205.

Henry, L., Wickham, H. and RStudio (2020). purrr: Functional Programming Tools.

Holdaway, S. and Wandsnider, L. (2008). *Time in Archaeology: Time Perspectivism Revisited*, University of Utah Press.
Kassambara, A. (2020). ggpubr: “ggplot2” Based Publication Ready Plots.
Kassambara, A. and Mundt, F. (2019). factoextra: Extract and Visualize the Results of Multivariate Data Analyses.
Knappett, C. (2021). 444Networks in Archaeology. In R. Light and J. Moody (eds.), *The Oxford Handbook of Social Networks*, Oxford University Press, p.0.
Martin, T., Zhang, X. and Newman, M. E. J. (2014). Localization and centrality in networks. *Phys. Rev. E* **90**: 052808.
Miller-Atkins, G. and Premo, L. S. (2018). Time-averaging and the spatial scale of regional cultural differentiation in archaeological assemblages. *STAR Sci. Technol. Archaeol. Res.* **4**: 12–27.
Mills, B. J. (2017). Social Network Analysis in Archaeology. *Annu. Rev. Anthropol.* **46**: 379–397.
Mills, B. J., Peeples, M. A., Aragon, L. D., Bellorado, B. A., Clark, J. J., Giomi, E. and Windes, T. C. (2018). Evaluating Chaco migration scenarios using dynamic social network analysis. *Antiquity* **92**: 922–939.
Mills, Barbara, Sudha Ram, Jeffery Clark, Scott Ortman, and Matthew Peeples. (2020). "cyberSW Version 1.0." Archaeology Southwest, Tucson.
Mullins, P. (2016). Webs of defense: Structure and meaning of defensive visibility networks in Prehispanic Peru. *J. Archaeol. Sci. Rep.* **8**: 346–355.
Östborn, P. and Gerding, H. (2014). Network analysis of archaeological data: a systematic approach. *J. Archaeol. Sci.* **46**: 75–88.
Pedersen, T. L. and RStudio (2021). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks.
Peeples, M.A. and Brughmans, T. (2023). Online Companion to Network Science in Archaeology. https://archnetworks.net, Accessed 2023-02-28.
Peeples, M. A., Mills, B. J., Clark, J. J., Aragon, L. D., Giomi, E. and Bellorado, B. A. (2016). Chaco Social Networks Database, Version 1.0.
Perreault, C. (2018). Time-Averaging Slows Down Rates of Change in the Archaeological Record. *J. Archaeol. Method Theory* **25**: 953–964.
Perreault, C. (2019). *The Quality of the Archaeological Record*, University of Chicago Press.
Prignano, L., Morer, I., Fulminante, F. and Lozano, S. (2019). Modelling terrestrial route networks to understand inter-polity interactions (southern Etruria, 950-500 BC). *J. Archaeol. Sci.* **105**: 46–58.
R Core Team (2022). R: A language and environment for statistical computing.
Ren, K. (2016). rlist: A Toolbox for Non-Tabular Data Manipulation.
Shaw, J. O., Coco, E., Wootton, K., Daems, D., Gillreath-Brown, A., Swain, A. and Dunne, J. A. (2021). Disentangling ecological and taphonomic signals in ancient food webs. *Paleobiology* **47**: 385–401.
Surovell, T. A. and Brantingham, P. J. (2007). A note on the use of temporal frequency distributions in studies of prehistoric demography. *J. Archaeol. Sci.* **34**: 1868–1877.
Surovell, T. A., Byrd Finley, J., Smith, G. M., Brantingham, P. J. and Kelly, R. (2009). Correcting temporal frequency distributions for taphonomic bias. *J. Archaeol. Sci.* **36**: 1715–1724.
Wickham, H. (2017). tidyverse: Easily Install and Load the “Tidyverse.”
Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H. and Dunnington, D. (2020). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.
Wickham, H., François, R., Henry, L. and Müller, K. (2021). dplyr: A Grammar of Data

Manipulation.
Wickham, H. and RStudio (2019). stringr: Simple, Consistent Wrappers for Common String Operations.
Wilke, C. O. (2019). cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”
Xie, Y., Cheng, J., Tan, X., Allaire, J. J., Girlich, M., Ellis, G. F., Rauh, J., htmlwidgets/lib), S. L. (DataTables in, htmlwidgets/lib), B. R. (selectize js in, htmlwidgets/lib), L. G. (noUiSlider in, htmlwidgets/lib), B. S. (jquery highlight js in, Pickering, A., RStudio and PBC (2021). DT: A Wrapper of the JavaScript Library “DataTables.”