

ORIGINAL ARTICLE

Measuring reciprocity: Double sampling, concordance, and network construction

Elsbeth Ready^{1,2*}  and Eleanor A. Power³ 

¹Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; ²Department of Anthropology, University of Florida, Gainesville, FL 32611, USA; and ³Department of Methodology, London School of Economics and Political Science, London WC2A 2AE, UK

*Corresponding author. Email: elsbeth_ready@eva.mpg.de

Action Editor: Ulrik Brandes

Abstract

Reciprocity—the mutual provisioning of support/goods—is a pervasive feature of social life. Directed networks provide a way to examine the structure of reciprocity in a community. However, measuring social networks involves assumptions about what relationships matter and how to elicit them, which may impact observed reciprocity. In particular, the practice of aggregating multiple sources of data on the same relationship (e.g., “double-sampled” data, where both the “giver” and “receiver” are asked to report on their relationship) may have pronounced impacts on network structure. To investigate these issues, we examine concordance (ties reported by both parties) and reciprocity in a set of directed, double-sampled social support networks. We find low concordance in people’s responses. Taking either the union (including any reported ties) or the intersection (including only concordant ties) of double-sampled relationships results in dramatically higher levels of reciprocity. Using multilevel exponential random graph models of social support networks from 75 villages in India, we show that these changes cannot be fully explained by the increase in the number of ties produced by layer aggregation. Respondents’ tendency to name the same people as both givers and receivers of support plays an important role, but this tendency varies across contexts and relationships type. We argue that no single method should necessarily be seen as the “correct” choice for aggregation of multiple sources of data on a single relationship type. Methods of aggregation should depend on the research question, the context, and the relationship in question.

Keywords: social networks; reciprocity; concordance; double sampling; informant accuracy

1. Introduction

Reciprocity—the mutual provisioning of support/goods—is recognized as a core feature of social relations in all of the social and behavioral sciences. Whether sociology (Gouldner, 1960), anthropology (Mauss, 1954), economics (Polanyi, 1957), or evolutionary biology (Trivers, 1971), reciprocity and exchange are seen as fundamental structuring forces of social groups. Dyadic reciprocity, where two partners mutually support each other, is a pervasive social norm dictating expected behavior in many contexts. Social scientists have also worked to identify reciprocity in practice, as an empirically observed behavioral pattern. However, understanding the importance of reciprocity in “real life,” and how it might vary across cultures and contexts, presents fundamental methodological challenges, both statistical and in terms of data collection. In the last two decades, developments in network science have begun to address the analytical challenge of reciprocity, namely the non-independence of reciprocal ties (Snijders, 2002; Safdari et al., 2021).

Here, we examine how common methods of network data collection and aggregation in the social sciences may impact the measurement of dyadic reciprocity.

1.1 Reciprocity

Classic anthropological work has highlighted the prevalence of a normative expectation of reciprocity cross-culturally, with many relationships understood as being characterized by mutual support and obligation, leading to long-term balanced exchange (Malinowski, 1922, 1926; Mauss, 1954; Sahlins, 1972). This has been reinforced by modeling work that demonstrates the many conditions under which reciprocity can be sustained (Trivers, 1971; Kranton, 1996). Quantitative empirical work has further demonstrated that reciprocity can explain much of the observed patterns of food sharing in humans and other primates (Jaeggi & Gurven, 2013). High levels of reciprocity have been found across a wide range of social and cultural contexts (Baldassarri, 2015; Gurven, 2004; Kasper & Borgerhoff Mulder, 2015; Power & Ready, 2018; Ready et al., 2020b; Vaquera & Kao, 2008; Szell et al., 2010). Such observations have convincingly established that reciprocity is pervasive across cultures, settings, and relationship types.

What we do not yet have a comprehensive understanding of, though, is how and why reciprocity varies across contexts, both cross-culturally and across domains such as friendship, food sharing, or lending. While contemporary methods of network analysis permit the contribution of reciprocity to tie formation to be estimated, the degree of reciprocity observed in empirical networks is also necessarily contingent on methods of data collection. A crucial barrier to such comparative work is differences in how data are collected and derived. Here, we outline how methodological decisions in survey-based research can impact resultant measures of reciprocity, hindering comparative efforts. We focus on two issues: informant recall and “double sampling.”

1.2 Informant accuracy and double sampling

While some network datasets are gathered through direct observation (e.g., records of phone calls), many are based on self-reports, because exchanges cannot always be directly observed and because some kinds of ties, such as feelings of closeness or friendship, are perceptual. This raises the issue of the accuracy of respondents' reporting of their relationships. Much work has been done to assess the accuracy of people's reports of their networks and has generally found it to be somewhat wanting (Marsden, 1990). A series of papers by Killworth et al. (Killworth & Bernard, 1976; Bernard & Killworth, 1977; Killworth & Bernard, 1979; Bernard et al., 1979, 1982), for example, damningly concluded that “informants are inaccurate. [...] [T]here appears to be systematic distortion of how informants recall just about everything” (Bernard et al., 1984). Inaccuracies in measurement are well recognized as impacting network structure and therefore network summary statistics (Marsden, 1990; Feld & Carter, 2002; Wang et al., 2012).

These “inaccuracies” to some extent reflect respondents' subjective perception of their networks, termed “cognitive social structures” (Krackhardt, 1987; Freeman et al., 1987). These cognitive schemas help people organize and remember their social worlds and are reflected in the responses they give to social support name generators. Subsequent work has therefore attempted to identify which relationships are likely to be forgotten or potentially falsely recalled, by which types of respondents. For example, respondents are more likely to name alters who are better connected (Marin, 2004) or of higher status (Grippa & Gloor, 2009; Ball & Newman, 2013; Shakya et al., 2017). Respondents may also be more likely to recall relationships that are strong (Marin, 2004) or that involve common partners (Brashears, 2013). Different types of relationships may also be more or less prone to biased reporting, for instance, people tend to remember reciprocity in “liking” relationships, and transitivity in influence relationships (De Soto, 1960). Certain types of relationships may be more likely to be remembered, and persons with many ties may be more likely to forget some of them (Bell et al., 2007). People's network positions and roles may also impact their perception of the network structure (Marineau et al., 2018; Ready et al., 2020a).

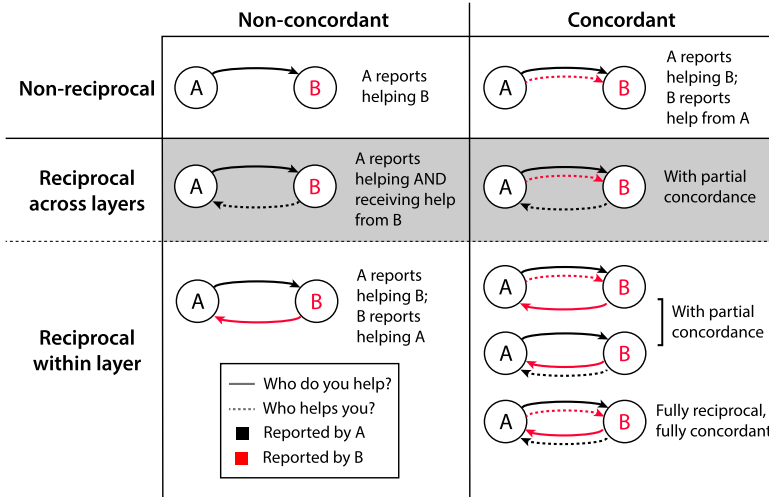


Figure 1. Reciprocity and concordance in directed networks. If a network is directed and double-sampled, assuming that A reports one tie with B, then there are eight possible graphs. Two of these possibilities produce reciprocity across layers, that is, reciprocal ties that would not exist if the data were single-sampled.

One way to estimate the “inaccuracies” in people’s nominations is to look at the agreement (concordance) between people nominally reporting on the same relationship. When there is complete sampling of a network, undirected ties or partnerships of various forms (e.g., friends, drug co-use partners, and sexual partners) should be reported by both members. However, studies have found far from perfect concordance when both parties are queried, with agreement generally ranging from 40% to 60% (Marsden, 1990), though some more recent work has found higher values (adams & Moody, 2007). It is crucial to note, however, that inconsistent reports do not necessarily reflect errors or inaccuracies in recall: they may legitimately reflect different perceptions of and investment in relationships (Carley & Krackhardt, 1996).

Past studies of concordance have generally focused on relationships that are presumed to be symmetric and so undirected (Marsden, 1990; adams & Moody, 2007).¹ In such cases, concordance and reciprocity are essentially equivalent. As such, these studies can speak only to reciprocity of *nomination*, not reciprocity of *exchange*. The latter is what is more generally of interest in the social sciences (beyond questions of measurement) and is what we concern ourselves with here.

In directed networks, reciprocity and concordance are not equivalent (Figure 1). Reciprocal ties occur within single-layer, directed networks when Ali says Beth gives to him, and Beth says Ali gives to her (Figure 1, bottom left). Concordant reports are generally not possible because individuals are not asked to report on the same set of ties (e.g., Ali and Beth each report who they give food to, but not who gives food to them). In some social network surveys, however, directed relationships are asked about in ways that facilitate a study of concordance between partners. A common practice with asymmetric relationships (i.e., ones that are inherently directed) is to “double-sample” (Nolin, 2008) by asking respondents not only who they turn to for various types of support but also who turns to them. Assuming complete sampling, this means that the same directed relationship should be reported on twice, once by the giver and once by the receiver.

Comparing network “layers” representing the same relationship in different directions (e.g., giving vs. receiving food) allows concordance to be measured in directed networks. Concordance occurs where both reporters agree on a tie in one direction and can be evaluated by comparing reports across double-sampled layers. For example, if Ali says he gives food to Beth, and Beth says Ali gives food to her, these are concordant reports for the tie A → B. Reciprocity occurs

where ties appear in both directions, which, as shown in Figure 1, can occur in multiple ways with double-sampled data. First, reciprocal ties can occur as they do in single-sampled networks (e.g., A reports $A \rightarrow B$ and B reports $B \rightarrow A$), irrespective of the aggregation of the two layers. Second, and uniquely for double-sampled data, reciprocal ties can also arise from the aggregation of the two layers, as when Ali reports that he both gives to and receives from Beth.

Double sampling is generally seen as a technique for dealing with the potential for bias and inaccuracy in people's reporting, with the assumption being that multiple insights into the same relationship will jointly be more accurate than a single report. This is similar to viewing each name generator as a layer of a multiplex network (Kivelä *et al.*, 2014). When relationships are reported on by multiple individuals, different approaches can be taken to deal with apparent inconsistencies in reporting. The simplest techniques involve considering the two networks in isolation (Koster, 2018; Simpson, *in press*), or combining them, either by taking the union (i.e., including all reported ties) (Nolin, 2010; Ready & Power, 2018) or, opposingly, taking the intersection (i.e., taking only concordant ties) (Krackhardt & Kilduff, 1990). These two decision rules have been referred to as allowing "unilateral nominations" for the union and requiring "mutual assent" for the intersection (Lee & Butts, 2018). Other techniques try to assess and then account for differences in each informant's accuracy, as with cultural consensus approaches (Romney & Weller, 1984; An & Schramski, 2015), while others use a more explicitly inferential and Bayesian approach (Butts, 2003; Lee & Butts, 2020; Newman, 2018; Young *et al.*, 2021; Redhead *et al.*, 2021). Aggregation techniques may also depend on whether networks are weighted or unweighted and how many sources of information are being integrated. Aggregation procedures will have important consequences for network density and possibly for other important network measures. Here, we consider the impact of simple aggregation techniques on the measurement of reciprocity.

1.3 The problem

If we wish to draw conclusions about observed differences in reciprocity across contexts—or even in similar contexts where different methods have been used—we need to ensure that we are making appropriate comparisons. However, comparison across different networks is a challenging problem (Faust & Skvoretz, 2002). With researchers taking potentially many different approaches when sampling, eliciting, and aggregating social support relationships, it is unclear how comparable measures of reciprocity calculated on different networks are. Biases in respondents' nominations—and researchers' attempts to account for those potential biases in nominations—may be impactful for calculations of reciprocity. In particular, insofar as reciprocal relationships are stronger, they may be inherently easier to recall. Further, if people have a desire for or an expectation of balanced, reciprocal relationships (Heider, 1958; Blau, 1964; Freeman, 1992; Krackhardt & Kilduff, 1999), then they may be particularly likely to report that the people they turn to for support similarly come to them.

In this paper, we ask to what extent measures of reciprocity are impacted by differences in the elicitation approach and aggregation technique used. We do so by drawing on a set of social support survey datasets that double-sampled relationships. These data allow us, first of all, to examine concordance in nominations across the various relationship types, and second, to consider how different decisions about how to treat non-concordant ties impact the resulting measures of reciprocity. Such work is a necessary precursor to any effort to compare measures of reciprocity from different networks, which may have been constructed in different ways.

2. Data

We primarily rely on social support network data from 75 villages in Karnataka, India (Banerjee *et al.*, 2013), with double sampling of four relationship types: borrowing and lending money, borrowing and lending household items, getting and giving advice, and visiting or hosting visitors.

These networks allow us to assess the variation across networks that were elicited using similar techniques in similar contexts. In this project, the social support network questions were asked of essentially all adult household members in a subsample of all households. Overall, roughly 46% of households and 25% of individuals (including children) per village were included. For our analyses, we consider networks of all surveyed individuals (i.e., we do not aggregate to households, and we exclude individuals who were not surveyed). Surveys were conducted by interviewers using pen and paper, with the survey form effectively limiting responses to four nominations for each question (though this limit was rarely reached by respondents).

For purposes of comparison, we also draw on several additional datasets that include double-sampled relationships, including two social support networks from two villages in Tamil Nadu, India (Power & Ready, 2019), country food and meal sharing networks from an Inuit village in the Canadian Arctic (Ready, 2016), and nine exchange and support networks from Mpimbwe, Tanzania (Kasper & Borgerhoff Mulder, 2015). These networks generally have higher coverage than the Karnataka data and provide an important counterpoint, given differences in the substance of the questions asked. The networks from the two villages in Tamil Nadu are quite comparable in substance and style to those from Karnataka, but with better coverage (94% and 97% of adult residents). The networks from the Canadian Arctic are based on interviews with household heads about common sharing partners at the household level, with 75% coverage of households. Finally, the networks from Tanzania are retrospective questions with concrete time horizons (e.g., “In the last seven days...”), asking about a range of exchange relationships, with 84% of adult household heads. These three datasets were gathered by interviewers using pen and paper surveys. All questions were free-list name generators where respondents could give as many nominations as they wished. The original survey prompts used in these studies are provided in Table S1 and summary statistics for the resulting networks are provided in Table S2.

Throughout the analyses, in order to distinguish between the two “layers” of double-sampled questions in a consistent manner, we use the term “incoming” layer to refer to name generators where the respondent reported receiving assistance or material aid (e.g., who lends you household items?). “Outgoing” layers refer to name generators where the respondent reported providing assistance or material aid to others (e.g., to whom do you lend household items?). “Ingoing” and “outgoing” therefore refer to the direction of the name generator from the perspective of the person answering the question. In key papers employing these datasets, the union of the double-sampled questions has generally been used as the network of interest (Banerjee et al., 2013; Ready & Power, 2018; Kasper & Borgerhoff Mulder, 2015).

3. Results

3.1 Concordance

We calculate concordance between all dyads, first for all ties reported across either the incoming or the outgoing network layers (Table 1 column “Overall” and Figure 2 show the distribution of values for the Karnataka dataset). We find that concordance across layers is generally low, with an average of 10% across all networks—decidedly lower than the 40%–60% range outlined in Marsden (1990) for a variety of undirected networks (e.g., “best friends” or “closest intimates”) or the more recent (undirected) social, sexual, and drug-use networks presented in Adams & Moody (2007), which are yet higher still. As shown in Figure 2, concordance is quite low in the Karnataka dataset, regardless of the relationship type.

Table 1 columns “Incoming” and “Outgoing” show the proportion of edges in each individual layer that are contributing to concordant ties in the combined set of layers. For instance, if Layer 1 had 20 edges, half of which were all present in Layer 2, and the Layer 2 had 100 edges, the values would be 9% (10/110) concordance overall across all unique reported edges, 50% (10/20) in Layer 1 and 10% (10/100) in Layer 2. The values reported in Table 1 demonstrate the extent to which

Table 1. Proportion of concordant ties in each double-sampled relationship in the datasets. “Incoming” refers to the name generator in which respondents reporting receiving a type of support, while “Outgoing” refers to respondents reporting giving support to others. The column “Overall” reports the proportion of concordant ties in the union of the incoming and outgoing layers

Network	Overall	Incoming	Outgoing
Karnataka (mean \pm sd)			
Money	0.09 \pm 0.03	0.15 \pm 0.05	0.17 \pm 0.05
Advice	0.11 \pm 0.04	0.18 \pm 0.07	0.22 \pm 0.08
Household items	0.10 \pm 0.03	0.18 \pm 0.05	0.19 \pm 0.05
Visit	0.10 \pm 0.03	0.18 \pm 0.05	0.19 \pm 0.05
Tamil Nadu, Aġakāpuram			
Loan	0.08	0.15	0.16
Household items	0.08	0.14	0.14
Tamil Nadu, Tenpattī			
Loan	0.08	0.14	0.15
Household items	0.06	0.12	0.11
Kangiqsujuaq			
Country food	0.20	0.40	0.28
Meals	0.23	0.43	0.33
Mtakula			
Childcare	0.18	0.29	0.31
Meals	0.11	0.17	0.23
Small loan	0.02	0.04	0.03
Large loan	0.08	0.12	0.18
Hospital	0.08	0.14	0.17
Labor	0.18	0.36	0.27
Clothes	0.08	0.14	0.17
Advice	0.08	0.11	0.26
Bicycle	0.07	0.11	0.14

the double-sampled layers capture different sets of edges: in none of the networks we examine can one of the layers be considered to be largely a subset of the other (the highest overlap is 43%, for the incoming meal sharing layer in Kangiqsujuaq). Generally, concordance in the outgoing layer is equal to or higher than that of the associated incoming layer, with the exception of the Kangiqsujuaq networks. This difference reflects the relative number of nominations in each of these layers: people often gave more nominations in response to questions asking from whom they received support than to questions about to whom they gave support.

The latter observation brings our attention to other differences in respondents’ nominations for incoming and outgoing prompts. In the Karnataka advice and loan networks, the incoming layers tend to have a more skewed degree distribution than the outgoing layers, with a few individuals receiving many nominations as providers of help. This is reflected in the heightened in-degree centralization for these layers (Figure 3), where centralization is defined as the tendency of a single point to be more central than other points in the network and the measure we use assesses the difference from the most centralized possible graph of the same size (i.e., a star) (Freeman, 1979). For certain relationships, when people are asked “who would you go to for support,” this is likely

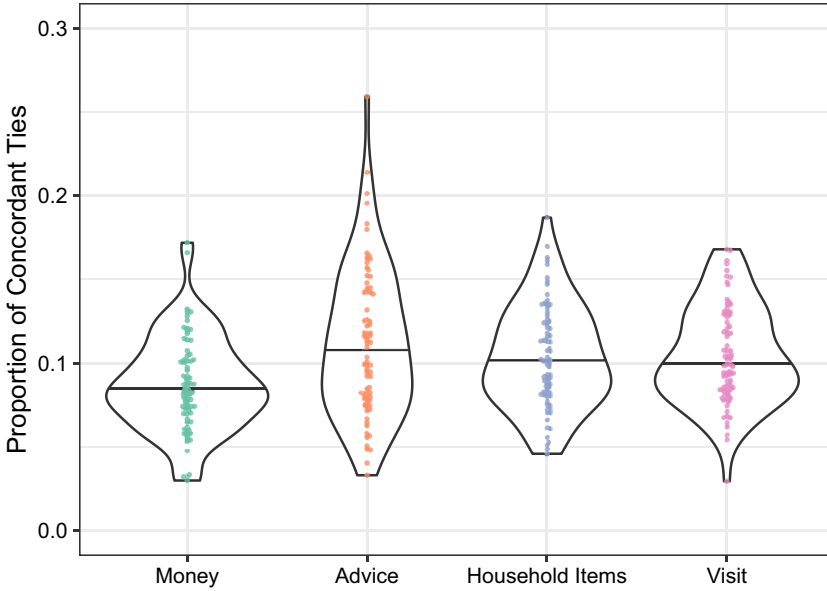


Figure 2. The proportion of concordant ties in each of the four double-sampled relationship types in the Karnataka dataset. Each point represents one of the 75 villages, and the violin plots show the overall probability densities with the median indicated by a line.

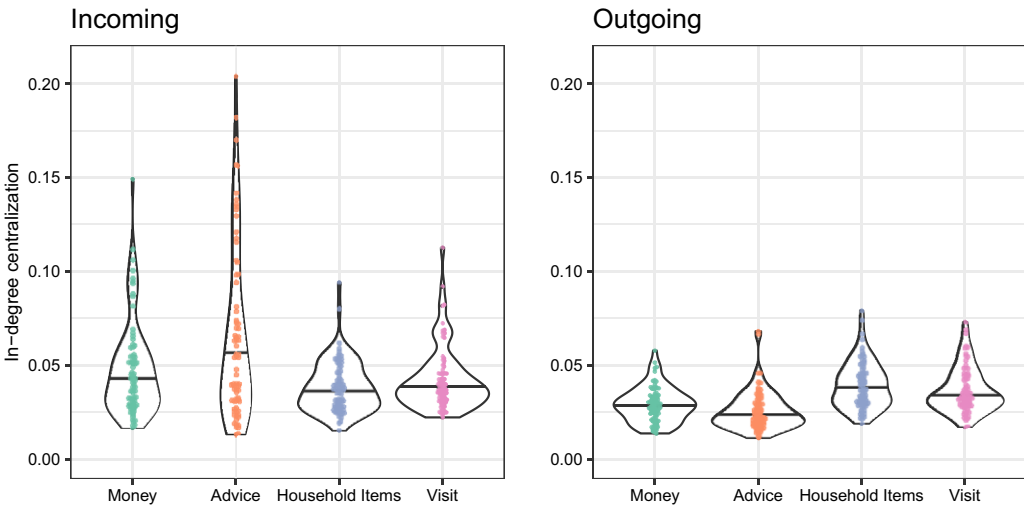


Figure 3. In-degree centralization of the incoming and outgoing Karnataka network layers. Each point represents one of the 75 villages, and the violin plots show the overall probability densities with the median indicated by a line.

to pick up a few key “hubs” or “stars” who have many ties. In contrast, when people are asked “who comes to you for support,” these structures are not detected because these high-activity people do not list substantially more alters. This phenomenon is also present in some of the other datasets, such as the Kangiqsujuaq food sharing data, where some households are consistently named by others as a source of food (centralization incoming layer = 0.26, outgoing layer = 0.12). This asymmetry appears to depend on the relationship type. For example, we do not see it across the incoming versus outgoing layers in the Karnataka household items and visitation networks,

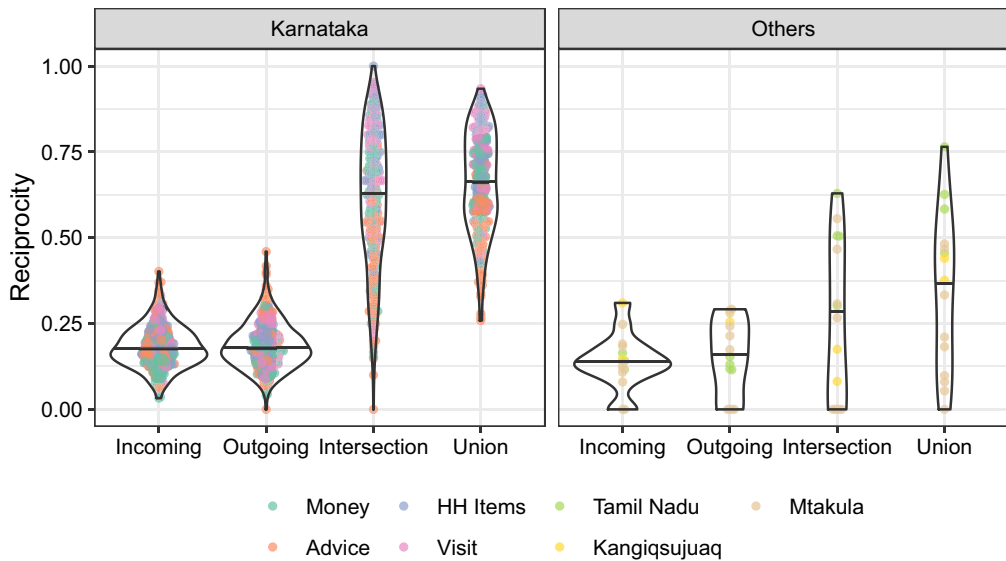


Figure 4. The proportion of reciprocal ties in the incoming and outgoing network layers, and in the aggregated intersection and union networks in the Karnataka dataset (left) and in the other datasets (right). Each point represents one of the 75 villages, and the violin plots show the overall probability densities with the median indicated by a line.

or the Kangiqsujuaq meal sharing data. Aggregation procedures will have an important impact on these asymmetries; in particular, they will be erased by considering only concordant ties.

3.2 Reciprocity

We calculate the proportion of reciprocal ties for each individual layer, as well as for the networks produced by the intersection and the union of the two layers (Figure 4 and Tables S2 and S3). Reciprocity is broadly comparable in outgoing and incoming layers (i.e., the reciprocity observed in giving a loan vs. asking for a loan are similar). However, when either the intersection or union is taken, we see a substantial increase in reciprocity in most of the networks.

Due to the low concordance in the datasets, taking the union results in networks with considerably more edges than either of the contributing layers and taking the intersection results in networks with substantially fewer edges. For example, in the Karnataka village networks, while the individual layers have an average density of 0.57%, the average density in the union networks is 1.03%, and the average density in the intersection networks is 0.11%. Despite the clear consequences of these aggregation procedures for the density of these merged networks, both show a dramatic increase in reciprocity. What might account for these increases?

We undertake further analyses to determine whether the observed changes in reciprocity are due primarily to the mathematics of aggregation. To do this, we fit multilevel exponential random graph models (Stewart & Schweinberger, 2018) to each layer of the double-sampled relationship types in the Karnataka dataset. We include terms for the number of edges, the number of nodes with in- and out-degree of zero through two (which encompasses the majority of nodes), reciprocity, and transitivity (geometrically weighted edgewise shared partnerships), to effectively match the core structure of each of the individual layers.² We then use the coefficients from the MLERGMs to simulate networks for each of the layers (e.g., borrow money and lend money) and generate aggregates of the two simulated layers, for each of the 75 villages, 100 times. These simulated aggregated networks provide a baseline against which to compare the observed aggregated networks.

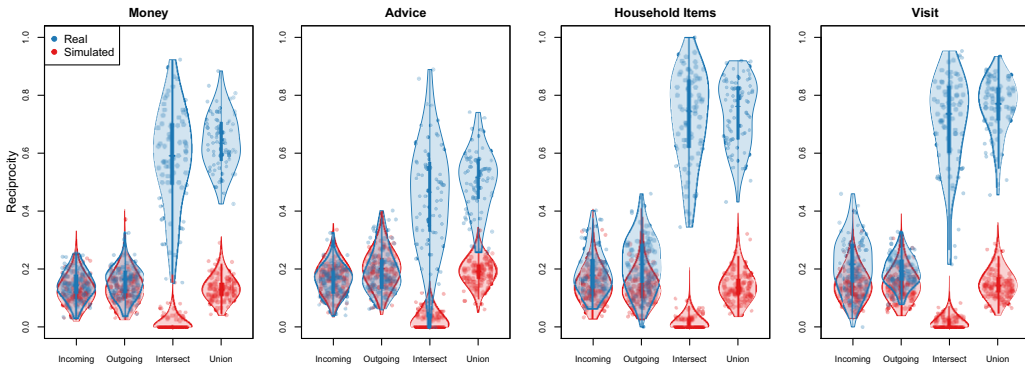


Figure 5. The effect of taking the union or intersection on reciprocity for real (blue) and simulated (red) networks for each of the four double-sampled relationship types in the Karnataka network dataset. Each point indicates one of the 75 villages, while the violin plots show the overall probability densities with the median indicated by a line and 50% quantiles indicated by the thicker midline. While the simulated networks match the levels of reciprocity in the original data reasonably well, when the intersection or union of the layers is taken, the level of reciprocity is drastically different from the observed networks.

In the case of the union networks, the increase in reciprocity could potentially be driven by the increase in density (i.e., more reciprocity could occur simply because there are more edges in the networks). Figure 5 shows reciprocity in the Karnataka dataset for the individual layers and aggregated networks, compared to reciprocity in the simulated networks. The results show that the dramatic increases in reciprocity observed in the empirical union networks is not reflected in the simulated union networks. Thus, the increase in reciprocity seen in the empirical union networks cannot simply be a consequence of increased network density.

It is informative to consider the source of the reciprocal ties in the union networks: reciprocal ties could occur either because the two parties report the directed relationship in a single layer, or because one party reports the tie across the two layers (Figure 1). In Table 2, we show the proportion of reciprocal ties in the union networks where ties are reported by both parties (“within-layer reciprocity,” Figure 1). Note that many entries are quite low, implying that much of the observed reciprocity in the union networks comes from a single reporter asserting a reciprocal relationship. However, this is highly variable across the datasets we examine. For instance, the networks from Kangiqsujuaq and Mtakula tend to have a much higher proportion of reciprocal ties that involve reports from both parties than the Indian networks (although many of the Mtakula networks are so sparse that the data are relatively uninformative).

The union network for advice in the Karnataka data has a lower overall level of reciprocity than the other networks in this dataset, but Table 2 suggests that when reciprocity is present in this network, it is somewhat more likely to have been reported by both parties. To explore this further, we calculate the probability of repeat naming in the Karnataka dataset (i.e., for all people listed by respondents in the first of two double-sampled questions, how many were listed again by the same respondent in the complementary question). These are often quite high: for example, sharing household items in Karnataka has a repeat nomination rate of 72% on average across the 75 villages. In the advice network, the rate of repeat nominations is low compared to other tie types in the same dataset (advice = 41%, visiting = 73%, and money = 58%) suggesting that people perceive advice relationships as being less reciprocal.

For the intersection networks, in comparison to the unaggregated networks, we see higher levels of reciprocity in the Karnataka networks despite *lower* density. Here again we use the simulated networks to provide a baseline for comparison. Starting from the simulated union networks, we randomly select the proportion of edges that were concordant in the associated empirical network and calculate reciprocity on the resulting subsetted network, effectively matching the empirical

Table 2. Proportion of reciprocal ties that are within layer in the union networks (i.e., where both respondents nominated each other in response to the same name generator)

Network	Proportion bilateral
Karnataka	
Money	0.18 ± 0.05
Advice	0.35 ± 0.09
Household items	0.19 ± 0.06
Visit	0.18 ± 0.06
Tamil Nadu, Aḷakāpuram	
Loan	0.19
Household items	0.15
Tamil Nadu, Teṅpaṭṭi	
Loan	0.25
Household items	0.16
Kangiṣṣujuaq	
Food	0.69
Meal	0.69
Mtakula	
Babysit	0.50
Meal	0.65
Small loan	0.83
Large loan	1.00
Hospital	0.00
Labor	0.40
Clothes	NA
Advice	0.55
Bicycle	0.67

intersection network size (Figure 5).³ Of the few edges that are retained in these simulated intersection networks, fewer still are reciprocal; the vast majority of the simulated networks have no reciprocal ties. This is what we would expect purely from the effect of the decrease in density, given the removal of edges. However, in the empirical data, reciprocity still increases substantially in the intersection networks, similar to the increase seen in the union networks. This suggests that concordant ties have a similar rate of reciprocity to non-concordant ties. The greater range of reciprocity values for the intersection networks reflects a heightened sensitivity to the particular set of edges that are retained, due to the small number of edges.

4. Discussion

Decisions on how to integrate multiple perspectives on what is nominally the same relationship have a dramatic effect on the structure of the resulting networks. Here, we have focused on reciprocity because of its theoretical and substantive importance, but other fundamental network measures also change through the process of aggregation. Perhaps most obviously, the density of

a network is impacted by the decision to include or censor different sets of edges. However, as we have shown in our analysis of the Karnataka networks, changes in network density are insufficient to explain the changes in reciprocity that we observe. Instead, the changes are linked to the interdependence between double-sampled name generators.

Contrary to previous work looking at concordance in undirected networks, here we have found that people's responses across questions about *directed* relationships show very low levels of concordance. The highest level of concordance in the set of networks we examined was 23%; most of the networks had levels closer to 10%. At first glance, this suggests remarkably low agreement on the existence of directed social ties. However, non-concordance does not necessarily mean that the relationships reported are not “real” or meaningful.

The interdependence of name generators is shaped by various factors including cognitive processes, aspects of the study design, and expectations for reciprocity in different kinds of relationships. First of all, conflicting reports between informants not only reflect potential informant inaccuracy (some ties may be forgotten or falsely reported) but also suggest that different people perceive their relationships differently (Krackhardt, 1987; Freeman et al., 1987; Comola & Fafchamps, 2014). For more subjective relationships that involve personal feelings and perceptions or hypothetical situations, givers and receivers may have different understandings of their relationship. While a person may feel that they *could* turn to another for advice, the potential advice-giver may not be aware that they are seen in this light. Even for more concrete relationships (like sharing food or household items), conflicting reports may reflect that a tie is differently valued by giver and receiver. Consider our finding that some prompts result in a few “stars” being named by many respondents. While “stars” may not list all of those who named them as provisioning support, this may *accurately* reflect that they perceive those relationships to be relatively unimportant, though they may be quite important to the receiver.

Different procedures for aggregating networks—taking the union (the most expansive set of edges) or taking the intersection (only concordant edges)—could imply conflicting presumptions about the “accuracy” of people's responses (and the sources of discrepancy between respondents), and about the dependence between name generators. For concrete prompts (e.g., “who did you lend money to in the past week?”), taking the union presumes that some respondents may be forgetful, but that all respondents are reporting “truthfully.” Taking the intersection presumes that some respondents may exaggerate, and that concordant ties are most likely to be “truthful.” Equally, if responses are seen as largely accurate but reflecting different subjective valuations of those relationships, then the union and intersection networks imply different analytical interests: the union seeks to capture the widest set of possible ties, however important, while the intersection focuses attention only on those that are mutually recognized. Given the low concordance in nominations we observe, these two sets of expectations result in very different networks, with the intersection comprising only a very small subset of reports.

Despite these differences, in the set of networks we have examined, the union and intersection procedures both generally result in levels of reciprocity that are substantially higher than what is observed in the individual layers. In the Karnataka network dataset, for example, both the union and the intersection result in reciprocity values that are roughly three times larger than what is seen in the individual layers. This is because respondents have a tendency to name the same people as both givers and receivers of exchanges (Table 2).

To some extent, this tendency to repeat nominations may be due to priming (i.e., naming the same people because they have already come to mind in the previous response). However, the extent of repeat nominations appears to depend not only on the study context but also on the type of relationship involved. In the Karnataka dataset, although the individual layers have similar levels of reciprocity across all relationship types, the increase in reciprocity in the advice networks after aggregation is less than that seen in the others (Figure 5). One potential explanation for this pattern is that certain relationship types may be more hierarchical and that people tolerate more “imbalance” in these relationships. For example, people may be more likely to ask

more senior persons for advice and may not expect to be asked for advice by them in return. The Karnataka networks indicate that people believe that many of their relationships are reciprocal (Appadurai, 1985), but also that they have differing expectations of reciprocity for different kinds of interactions (such as providing advice vs. lending household items). Although it is not clear from these data the extent to which these differences reflect actual behavior, it is clear that double-sampled relationships are potentially revealing of people's *expectations* about what kinds of relationships should be reciprocal.

We have focused on how network elicitation and aggregation might impact observed reciprocity, but we are leaving questions about why these processes impact network structure largely unanswered (e.g., due to priming, due to expectations of reciprocity, etc.). To some extent, this is due to the inherent ambiguity of the data that is being elicited. For instance, we have suggested that it is the nature of the double-sampled questions that is leading to potentially inflated reciprocity values, but it is also possible that “actual” reciprocity makes the reporting of concordant ties more likely. In the Karnataka data, the latter explanation cannot fully account for the patterns we have observed (because we do not see higher levels of reciprocity in the intersection than in the union networks), but it is worth emphasizing that the complexities and subjectivities of the data we are considering makes it particularly difficult to adjudicate between alternative explanations.

With the multilevel exponential random graph models, we attempted to identify some of the potential mechanisms that could account for the seemingly inflated reciprocity values observed in the union and intersection networks. Much more could potentially be done with such models to try to more fully identify the potential mechanisms at play. We have already suggested that some of the patterns observed may be driven by individuals with high in-degree, which could potentially be modeled with receiver effect terms. Future work should develop generative models aimed at better capturing the processes that produce the interdependence or apparent disagreement across layers and respondents. This may be through assessing the “credibility” of reporters (An & Schramski, 2015) or by using Bayesian approaches to assess consensus or disagreement about the existence or strength of particular ties (Butts, 2003; Lee & Butts, 2020; Newman, 2018; Young et al., 2021; Redhead et al., 2021).

5. Conclusion

Double sampling is generally seen as being of value insofar as it will help to create more complete and more accurate representations of people's social connections, thanks to the multiple perspectives it provides. Here, we have challenged this notion by showing that different aggregation techniques can result in very different network structures. What is clear from our analysis is that double-sampled reports are not independent, and that double sampling has the potential to reveal important insights into how people understand their relationships. So, what should researchers consider when designing studies that gather or use directed (and potentially double-sampled) network data?

First, are you concerned with actual interactions or with people's perception of their networks? As emphasized in the cognitive networks literature (Krackhardt, 1987; Freeman et al., 1987; Carley & Krackhardt, 1996; Marin, 2004), people's beliefs about their relationships, even if they are inaccurate representations of “actual” interactions, can be highly relevant to understanding their behavior. In these circumstances, non-concordance may not be of particular concern, as each respondent's nominations can potentially be treated as an accurate representation of their own perception. However, in such cases, it is important to reflect on what the aggregate of people's perceptions actually represents, and how it may (or may not) have real-world consequences (Krackhardt, 1987; Freeman et al., 1987).

Second, what are the local norms of reciprocity for the tie type of interest? As we have suggested here, strong norms of reciprocity may lead people to “inflate” their responses by repeatedly

naming individuals. This can result in drastically increased reciprocity in aggregated networks. This should be especially true for relationship types that are expected to be mutually supportive and balanced, while relationships that are expected to be hierarchical and imbalanced may be less impacted. Again, however, the “inflation” here is not necessarily an inaccuracy. For instance, a person who asks a favor once a year may report a reciprocal relationship that is true but not remembered by a partner who is much more active. Close attention also needs to be paid to the impacts of question order and priming on the repeat naming of alters in network questionnaires.

Third, how concentrated is the resource being exchanged or the activity undertaken? The presence of “star” structures, where a few individuals have a large number of ties, may often be linked to an uneven distribution of resources between individuals (e.g., there may be relatively few money lenders in a community). Relationships that are more constrained by time or physical distance, such as visiting or borrowing everyday items, may be less likely to show such asymmetries. If such stars are not anticipated but appear in the data, this may reflect potential biases in people’s responses, such as a potential bias in naming patterns related to prestige or reknown (e.g., if elected officials are often named). Indeed, previous research suggests that a low ratio of self-reported to alter-reported interactions may be linked to leadership, with perceived leaders (even informal leaders) being more likely to underreport their activity (Grippa & Gloor, 2009). We encourage network researchers to reflect on the relationships between network structure, other contextual factors such as resource distribution, and responses in social network surveys. Asymmetries revealed through double sampling may shed light onto these relationships and can potentially be used to inform the choice of aggregation technique.

Recent research has suggested that, between the union and the intersection, the intersection may provide more accurate depictions of social structure, particularly in sparse graphs (Lee & Butts, 2018). But in datasets like those considered here, this could mean discarding up to 90% of the data. Given the strong presumptions inherent in the union and the intersection, other more nuanced aggregation techniques should clearly be considered. Above, we suggested that more complex models have the potential to help researchers better understand the data-generating process and the potential biases within it; such models may then also be useful in helping researchers account for these biases to arrive at a representation of the network that is a better basis for inference (Redhead et al., 2021). However, given the subjectivity involved in the perception of relationships, it is not the case that aggregation methods, whether simple or complex, can always recover an objective “groundtruth.” There is not a “one size fits all” solution here: the researcher must consider the particular context and relationships in question in order to determine which approach to aggregation is most principled.

The need to consider the particularities of each network further highlights that comparisons *across* networks need to be undertaken with caution. Slightly different prompts may produce very different networks (e.g., “who would you ask for advice?” vs. “who did you ask for advice in the last week?”). As we have shown here, comparing single layers to aggregated networks may be particularly misleading. Even within networks aggregated in the same way, different norms may lead to more “inflation” of reciprocity with double sampling. Researchers should pay close attention to the sociocultural setting, data collection approaches (including the elicitation technique such as name generators vs. roster methods, the particular prompt, question order, etc.), and data aggregation techniques, in order to determine if direct comparison across datasets is warranted. We note that the issues discussed here should also be considered when aggregating multiple relationship types into a single network (e.g., different kinds of exchanges aggregated to a single “social support” network).

Overall, then, we encourage researchers to think carefully about how respondents will engage with the questions that are posed to them and how this will impact the observed network(s) under study. Research designs can help assess (within-study) measurement validity and contribute to a better understanding of some of the issues raised in this paper. Such efforts should be undertaken early, at the stage of designing and piloting survey instruments. As one of the reviewers suggested,

using multidimensional scales to assess relationship strength may provide more detailed data that can hopefully help tease out the reasons for lack of concordance between respondents. For instance, in response to a question such as “who do you ask for advice?” one might consider a scaled answer such as “could be asked if needed,” “have previously asked,” and “asked within the last month.” One would expect concordance to correlate with such a scale.

In conclusion, we see great value in double sampling, despite its complexities. We emphasize that there is no one “right” way to collect and aggregate data on people’s relationships. Rather, appropriate methods to represent a network depend on the research question and require careful reflection about the properties of ties that are considered to be important to the outcome of interest. What appear to be methodological decisions are thus also theoretical ones.

Acknowledgments. The authors would like to thank members of the Department of Human Behavior, Ecology and Culture at the Max Planck Institute for Evolutionary Anthropology for helpful comments on the paper and Monique Borgerhoff Mulder for access to the Mtakula dataset.

Competing interests. None.

Funding statement. Data collection by the authors was supported by National Science Foundation grants PLR-1303874 and IBSS-143019, and by Social Sciences and Humanities Research Council of Canada grant 752-2010-1089.

Data availability statement. The Karnataka dataset is available at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/21538>. Code for our analyses and additional information on accessing the datasets can be found at https://github.com/elspethr/measuring_reciprocity.

Supplementary materials. For supplementary material for this article, please visit <http://dx.doi.org/10.1017/nws.2021.18>

Notes

1 A different sampling approach, used in “cognitive social structure” studies (Krackhardt, 1987), asks group members to report on all possible relationships within the group. We do not consider such sampling approaches here, but instead refer readers to the work on ways to aggregate such accounts into various “locally aggregated structures” (Lee & Butts, 2018).

2 We note that we use MLERGMs for this analysis because other readily available simulation methods, such as the standard configuration model, do not generate networks that are similar to the empirical data. In particular, they underestimate reciprocity and transitivity in the single-layer networks. We felt that preserving these properties in the simulated data was required in order to answer our research question.

3 We note that we derive the intersection this way, rather than taking the actual intersection from two simulated networks (as we do with the union), because it results in graphs with very few or no edges.

References

- adams, j., & Moody, J. (2007). To tell the truth: Measuring concordance in multiply reported network data. *Social Networks*, 29(1), 44–58.
- An, W., & Schramski, S. (2015). Analysis of contested reports in exchange networks based on actors credibility. *Social Networks*, 40, 25–33.
- Appadurai, A. (1985). Gratitude as a social mode in South India. *Ethos*, 13(3), 236–245.
- Baldassarri, D. (2015). Cooperative networks: Altruism, group solidarity, reciprocity, and sanctioning in Ugandan producer organizations. *American Journal of Sociology*, 121(2), 355–395.
- Ball, B., & Newman, M. E. J. (2013). Friendship networks and social status. *Network Science*, 1(1), 16–30.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144), 1236498–1–7.
- Bell, D. C., Belli-McQueen, B., & Haider, Ali. (2007). Partner naming and forgetting: Recall of network members. *Social Networks*, 29(2), 279–299.
- Bernard, H. R., & Killworth, P. D. (1977). Informant accuracy in social network data II. *Human Communication Research*, 4(1), 3–18.
- Bernard, H. R., Killworth, P. D., Kronenfeld, D., & Sailer, L. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology*, 13, 495–517.
- Bernard, H. R., Killworth, P. D., & Sailer, L. (1979). Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, 2(3), 191–218.

- Bernard, H. R., Killworth, P. D., & Sailer, L. (1982). Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data. *Social Science Research*, 11(1), 30–66.
- Blau, P. M. (1964). *Exchange and power in social life*. New York: J. Wiley.
- Brashears, M. E. (2013). Humans use compression heuristics to improve the recall of social networks. *Scientific Reports*, 3, 1513.
- Butts, C. T. (2003). Network inference, error, and informant (in)accuracy: A Bayesian approach. *Social Networks*, 25(2), 103–140.
- Carley, K. M., & Krackhardt, D. (1996). Cognitive inconsistencies and non-symmetric friendship. *Social Networks*, 18(1), 1–27.
- Comola, M., & Fafchamps, M. (2014). Testing unilateral and bilateral link formation. *The Economic Journal*, 124(579), 954–976. Publisher: Wiley.
- De Soto, C. B. (1960). Learning a social structure. *The Journal of Abnormal and Social Psychology*, 60(3), 417.
- Faust, K., & Skvoretz, J. (2002). Comparing networks across space and time, size and species. *Sociological Methodology*, 32(1), 267–299.
- Feld, S. L., & Carter, W. C. (2002). Detecting measurement bias in respondent reports of personal networks. *Social Networks*, 24(4), 365–383.
- Freeman, L. C. (1992). Filling in the blanks: A theory of cognitive categories and the structure of social affiliation. *Social Psychology Quarterly*, 55(2), 118–127.
- Freeman, L. C., Romney, A. K., & Freeman, S. C. (1987). Cognitive structure and informant accuracy. *American Anthropologist*, 89(2), 310–325.
- Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3), 215–239.
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25(2), 161–178.
- Grippa, F., & Gloor, P. A. (2009). You are who remembers you. Detecting leadership through accuracy of recall. *Social Networks*, 31(4), 255–261.
- Gurven, M. (2004). Reciprocal altruism and food sharing decisions among Hiwi and Ache hunter-gatherers. *Behavioral Ecology and Sociobiology*, 56(4), 366–380.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Jaeggi, A. V., & Gurven, M. (2013). Reciprocity explains food sharing in humans and other primates independent of kin selection and tolerated scrounging: A phylogenetic meta-analysis. *Proceedings of the Royal Society B: Biological Sciences*, 280(1768), 20131615.
- Kasper, C., & Borgerhoff Mulder, M. (2015). Who helps and why?: Cooperative networks in Mpimbwe. *Current Anthropology*, 56(5), 701–732.
- Killworth, P. D., & Bernard, H. R. (1976). Informant accuracy in social network data. *Human Organization*, 35(3), 269–286.
- Killworth, P. D., & Bernard, H. R. (1979). Informant accuracy in social network data III: A comparison of triadic structure in behavioral and cognitive data. *Social Networks*, 2(1), 19–46.
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3), 203–271.
- Koster, J. (2018). Family ties: The multilevel effects of households and kinship on the networks of individuals. *Open Science*, 5(4), 172159.
- Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9(2), 109–134.
- Krackhardt, D., & Kilduff, M. (1990). Friendship patterns and culture: The control of organizational diversity. *American Anthropologist*, 92(1), 142–154.
- Krackhardt, D., & Kilduff, M. (1999). Whether close or far: Social distance effects on perceived balance in friendship networks. *Journal of Personality and Social Psychology*, 76(5), 770–782.
- Kranton, R. E. (1996). Reciprocal exchange: A self-sustaining system. *The American Economic Review*, 86(4), 830–851.
- Lee, F., & Butts, C. T. (2018). Mutual assent or unilateral nomination? A performance comparison of intersection and union rules for integrating self-reports of social relationships. *Social Networks*, 55, 55–62.
- Lee, F., & Butts, C. T. (2020). On the validity of perceived social structure. *Journal of Mathematical Psychology*, 98, 102384.
- Malinowski, B. (1922). *Argonauts of the western Pacific: An account of native enterprise and adventure in the archipelagoes of Melanesian New Guinea*. London: G. Routledge & Sons, Ltd.
- Malinowski, B. (1926). *Crime and custom in savage society*. London: K. Paul, Trench, Trubner & Co, Ltd.
- Marin, A. (2004). Are respondents more likely to list alters with certain characteristics?: Implications for name generator data. *Social Networks*, 26(4), 289–307.
- Marineau, J. E., Labianca, G., Brass, D. J., Borgatti, S. P., & Vecchi, P. (2018). Individuals' power and their social network accuracy: A situated cognition perspective. *Social Networks*, 54, 145–161.
- Marsden, P. V. (1990). Network data and measurement. *Annual Review of Sociology*, 16(1), 435–463.
- Mauss, M. (1954). *The gift: Forms and functions of exchange in archaic societies*. Glencoe, IL: Free Press.
- Newman, M. E. J. (2018). Network structure from rich but noisy data. *Nature Physics*, 14(6), 542–545.

- Nolin, D. A. (2008). *Food-sharing networks in Lamalera, Indonesia: Tests of adaptive hypotheses*. Ph.D. dissertation, University of Washington, Seattle.
- Nolin, D. A. (2010). Food-sharing networks in Lamalera, Indonesia: Reciprocity, kinship, and distance. *Human Nature*, 21(3), 243–268.
- Polanyi, K. (1957). The economy as instituted process. In K. Polanyi, C. Arensberg, & H. Pearson (Eds.), *Trade and market in the early empires: Economies in history and theory* (pp. 243–270). New York: Free Press.
- Power, E. A., & Ready, E. (2018). Building bigness: Reputation, prominence, and social capital in rural South India. *American Anthropologist*, 120(3), 444–459.
- Power, E. A., & Ready, E. (2019). Cooperation beyond consanguinity: Post-marital residence, delineations of kin, and social support among South Indian Tamils. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1780), 20180070.
- Ready, E. (2016). *Food, sharing, and social structure in an Arctic mixed economy*. Ph.D. dissertation, Stanford University, Stanford.
- Ready, E., Habecker, P., Abadie, R., Davila, C. A., Rivera Villegas, A., Khan, B., & Dombrowski, K. (2020a). Comparing social network structures generated through sociometric and ethnographic methods. *Field Methods*, 32(4), 416–432.
- Ready, E., Habecker, P., Abadie, R., Khan, B., & Dombrowski, K. (2020b). Competing forces of withdrawal and disease avoidance in the risk networks of people who inject drugs. *PLoS ONE*, 15(6), e0235124.
- Ready, E., & Power, E. A. (2018). Why wage earners hunt: Food sharing, social structure, and influence in an Arctic mixed economy. *Current Anthropology*, 59(1), 74–97.
- Redhead, D., McElreath, R., & Ross, C. T. (2021). Reliable network inference from unreliable data: A tutorial on latent network modeling using STRAND. Psyarxiv, 10.31234/osf.io/mkp2y.
- Romney, A. K., & Weller, S. C. (1984). Predicting informant accuracy from patterns of recall among individuals. *Social Networks*, 6(1), 59–77.
- Safdari, H., Contisciani, M., & De Bacco, C. (2021). A generative model for reciprocity and community detection in networks. arXiv, 2012.08215.
- Sahlins, M. D. (1972). *Stone Age economics*. Chicago: Aldine-Atherton.
- Shakya, H. B., Christakis, N. A., & Fowler, J. H. (2017). An exploratory comparison of name generator content: Data from rural India. *Social Networks*, 48, 157–168.
- Simpson, C. R. (in press). On the structural equivalence of coresidents and the measurement of village social structure. *Social Networks*.
- Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1–40.
- Stewart, J., & Schweinberger, M. (2018). *mlergm: Multilevel exponential-family random graph models*. R package version 0.1.
- Szell, M., Lambiotte, R., & Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31), 13636–13641.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35–57.
- Vaquera, E., & Kao, G. (2008). Do you like me as much as I like you? Friendship reciprocity and its effects on school outcomes among adolescents. *Social Science Research*, 37(1), 55–72.
- Wang, D. J., Shi, X., McFarland, D. A., & Leskovec, J. (2012). Measurement error in network data: A re-classification. *Social Networks*, 34(4), 396–409.
- Young, J.-G., Cantwell, G. T., & Newman, M. E. J. (2021). Bayesian inference of network structure from unreliable data. *Journal of Complex Networks*, 8, cnaa046.

Cite this article: Ready E and Power E.A. Measuring reciprocity: Double sampling, concordance, and network construction. *Network Science* <https://doi.org/10.1017/nws.2021.18>