Journal of the Royal Statistical Society Series A: Statistics in Society, 2022, **00**, 1–21
https://doi.org/10.1093/jrsssa/qnac004

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Original Article

Latent network models to account for noisy, multiply reported social network data

Caterina De Bacco[1], Martina Contisciani[1], Jonathan Cardoso-Silva[2,3], Hadiseh Safdari[1], Gabriela Lima Borges[4], Diego Baptista[1], Tracy Sweet[5], Jean-Gabriel Young[6], Jeremy Koster[7,8], Cody T. Ross[4], Richard McElreath[4], Daniel Redhead[4] and Eleanor A. Power[2,9]

[1]Cyber Valley, Max Planck Institute for Intelligent Systems, Tuebingen, Germany
[2]Department of Methodology, London School of Economics and Political Science, London, UK
[3]Data Science Institute, London School of Economics and Political Science, London, UK
[4]Department of Human Behaviour, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[5]Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, USA
[6]Department of Mathematics and Statistics and Vermont Complex Systems Center, University of Vermont, Burlington, VT, USA
[7]Department of Anthropology, University of Cincinnati, Cincinnati, OH, USA
[8]Division of Behavioral and Cognitive Sciences, National Science Foundation, Alexandria, VA, USA
[9]Santa Fe Institute, Santa Fe, NM, USA

Address for correspondence: Eleanor A. Power, Department of Methodology, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Email: e.a.power@lse.ac.uk

Abstract

Social network data are often constructed by incorporating reports from multiple individuals. However, it is not obvious how to reconcile discordant responses from individuals. There may be particular risks with multiply reported data if people's responses reflect normative expectations—such as an expectation of balanced, reciprocal relationships. Here, we propose a probabilistic model that incorporates ties reported by multiple individuals to estimate the unobserved network structure. In addition to estimating a parameter for each reporter that is related to their tendency of over- or under-reporting relationships, the model explicitly incorporates a term for 'mutuality', the tendency to report ties in both directions involving the same alter. Our model's algorithmic implementation is based on variational inference, which makes it efficient and scalable to large systems. We apply our model to data from a Nicaraguan community collected with a roster-based design and 75 Indian villages collected with a name-generator design. We observe strong evidence of 'mutuality' in both datasets, and find that this value varies by relationship type. Consequently, our model estimates networks with reciprocity values that are substantially different than those resulting from standard deterministic aggregation approaches, demonstrating the need to consider such issues when gathering, constructing, and analysing survey-based network data.

Keywords: Social network data, mutuality, reliability, variational inference, latent network, network measurement

1 Introduction

Social network analysis has emerged as a fruitful framework for social scientists to represent and understand social relationships and their consequences (Borgatti et al., 2009). For example, patterns of interaction among people, as well as peoples' perceptions of their relationships, have been found to be important for their material wealth (Jackson, 2021), social position and

Received: December 21, 2021. **Revised:** July 27, 2022. **Accepted:** August 17, 2022

welfare (Lin, 2002; Redhead & Power, 2022), and health and well-being (Holt-Lunstad et al., 2015; Perkins et al., 2015).

While new data sources now allow for the study of digitally mediated interactions (such as social media, mobile phone records, and other trace data; Eagle et al., 2009; Lazer et al., 2021; Park et al., 2018), social scientists' interest in day-to-day interactions and interpersonal relations are not always amenable to direct observation. Researchers, therefore, continue to rely on surveys where respondents identify the people with whom they have interactions or social relationships (Burt, 1984). A variety of approaches exist for eliciting self-reported network ties from respondents. Most common is the 'name generator' method, where respondents are asked to list the names of those with whom they have different types of relationships or interactions. Other approaches require a full roster, where respondents are asked about their relationship(s) with a set of possible partners (Marsden, 2005; Ross & Redhead, 2022; Warner et al., 1979).

Importantly, survey-based elicitations can be used not only for accounts of concrete interactions or exchanges, but can also facilitate a representation of respondents' subjective perceptions of their connections (Freeman, 1992; Krackhardt, 1987). Questions may be framed around more qualitative sentiments towards others—such as in friendships—and so do not merely document concrete interactions or observed events of exchange. For many substantive research questions, an individual's imperfect perception of their social relationships may be as (if not more) important as observable events of interaction or exchange. This has been highlighted by empirical research suggesting that individuals place considerable weight on their subjective relationships when making important decisions about who to cooperate with or support (Power, 2017; Redhead & von Rueden, 2021; von Rueden et al., 2019), and by work demonstrating that such relationships have strong associations with many important social and health-related outcomes (Kristiansen, 2004; Smith & Christakis, 2008).

The applicability of self-reported network data, however, has been subject to enduring debate within the social networks literature. Particularly when prompts query concrete exchanges or interactions, the quality of such data rests on the reliability of the self-reports that respondents provide, and numerous empirical studies have highlighted a plethora of potential biases in responses (Bernard et al., 1984; Killworth & Bernard, 1976). There is evidence that respondents' recall of their ties can be low, even over short periods of time (Brewer, 2000). For example, women within two West African communities were only able to accurately recall between 53% and 59% of their interactions across a 24-hr period prior to surveying (Adams et al., 2006). Alongside this, individual differences in the ability to recall ties may be predicted by relationship type, the number of partners a person has, and the duration of a given relationship (An, 2022; Bell et al., 2007). Both theoretical studies and empirically observed patterns of nominations suggest that individuals expressing particular attributes (e.g., high social status or power; Simpson et al., 2011) are more readily named, regardless of whether a relationship actually exists (Ball & Newman, 2013; Marin, 2004; Marineau et al., 2018; Redhead et al., Accepted; Shakya et al., 2017). The order in which questions appear within a survey, and the mode of elicitation, may further influence responses (Eagle & Proeschold-Bell, 2015; Pustejovsky and Spillane, 2009). That is, respondents have been shown to become fatigued, and report fewer relationships, when asked several name generator questions (Yousefi-Nooraie et al., 2019). Responses can also vary between interviewers, based in part on their attributes and their dynamic with the interviewee (Lungeanu et al., 2021; Marsden, 2003).

Noting all of these potential biases, one common practice is to obtain multiple reports on any single tie within a network. For relationships that are understood to be undirected, this is inherently captured with a single name generator question (i.e., both members of a friendship have the opportunity to report it). Previous research has found mixed results as to the concordance between respondents about the existence of their social relationships, with agreement in nominations ranging between 40% and 90% (Adams & Moody, 2007; Marsden, 1990). For relationships that are understood to be directed, multiple queries are necessary. One common approach is to 'double sample' a relationship, by asking respondents both who they go to for some type of assistance, and also who comes to them (Nolin, 2008). When combined with complete censusing of individuals, double sampling provides two perspectives on all relationships within a network, as both the giver and receiver have an opportunity to name their partner in each prompt. A recent survey of double-sampled network data has suggested that concordance between reporters is low, with an overall average of 10% agreement (Ready & Power, 2021).

Respondents need not be limited to reporting on the relationships in which they are directly involved, but may also be asked about the relationships between other individuals within the network. This type of data has been collected through 'cognitive social structures' roster designs—where respondents report on the relationships between all individuals within the network (Krackhardt, 1987; Newcomb, 1961)—though respondent fatigue means that this elicitation technique is somewhat uncommon. When it has been used, it has also shown relatively low levels of concordance between responses, highlighting that individual differences may guide respondents' perceptions of their own relationships and the relationships of others (see Brands, 2013, for a review).

Low levels of inter-respondent concordance suggest that while having multiple reports on any relationship certainly provides new information, it does not necessarily resolve the issue of bias in reporting. Indeed, new issues may be introduced, if there are, for example, different reporting propensities for different queries. One key issue for double-sampled data, in particular, may be people's expectation of, or desire for, mutually supportive, balanced relationships (Heider, 1958). The use of multiple prompts entailed in double sampling may lead to an inflation of apparent reciprocity, driven primarily by people's propensity to name the same individuals across both prompts (Ready & Power, 2021). We use the term 'mutuality' here to refer to this apparent inflation of reciprocity. Overall, the low level of concordance found in multiply reported data raises the question of how to statistically account for the ambiguity introduced by conflicting reports of the same potential tie.

To examine the individual biases that shape self-reports of ties, and to estimate the effect of mutuality on the core properties of a network, we introduce a new latent network model for directed ties that is able to combine multiply reported network data, while accounting for the variable 'reliability' of respondents. Thus, we estimate a latent network, where the probability of an unobserved tie between two nodes is jointly dependent on the reports of multiple individuals and the reliability of those individuals. We validate our model by simulating noisy reports from a true network of ties, and then verify that we are able to recover the true generative network and the individual-level reporter reliability and mutuality parameters. Finally, we evaluate our model using two empirical datasets that feature double-sampled questions, one based on a 'name generator' design and the other based on a roster method design. We conclude by discussing our findings and outlining possible extensions of the model.

1.1 Related work

In the social sciences, simple deterministic rules are often used to aggregate multiple reports on what should nominally be the same relationship (Krackhardt, 1987; Lee & Butts, 2018). When data are collected via double sampling, for example, it is sometimes assumed that if one party forgets to report a relationship when asked (e.g., when they are asked who they give advice to), the other party may report that tie (e.g., when they are asked who they receive advice from). With such an expectation, the union of the two name generators is typically used (e.g., Nolin, 2010; Ready & Power, 2018). Alternatively, it could be assumed that relationships are only salient when they are mutually recognized; under such an expectation, the intersection of the two name generators would be preferred (e.g., Krackhardt & Kilduff, 1990). These aggregation rules rely on simple but strong expectations and presume consistency in how reporters respond to these questions. This, paired with the fact that the statistical tools used most frequently in the social sciences (e.g., exponential random graph models; Robins et al., 2007) assume that reported ties are a 'true' representation of a given network, can potentially lead to serious misrepresentations in the social relations of interest in a given study.

Several statistical methods have been proposed to resolve discordant reports for social network analysis (Butts, 2003; Holland et al., 1983; Kenny & La Voie, 1984; Killworth & Bernard, 1976; Redhead et al., Accepted; Sewell, 2019; Sosa & Rodríguez, 2021). Similar methods have also been introduced in other fields, like systems engineering (Amini et al., 2004), the biological sciences (D'haeseleer & Church, 2004; Hobson et al., 2021; Sprinzak et al., 2003), and physics (Newman, 2018b). Recently, for example, social scientists have attempted to tackle the problem of concordance by computing a 'credibility score' for every individual within a network, and determining whether a given tie exists based on each reporter's assigned credibility (An & Schramski, 2015).

Considering this broad literature, we focus on methods most similar to our own, namely approaches that rely on an explicit generative model for reports that provide only imperfect information about a true network of ties. For cognitive social structure data, in which each person reports on ties between every pair of people in the network, both Sewell (2019) and Sosa and Rodríguez (2021) have introduced models that aggregate network tie information across all reporters and simultaneously estimate error parameters for each reporter. The model proposed by Butts (2003) is more similar to our work in that it accommodates fewer reports on each tie and assumes the existence of a true underlying network. More recently, Redhead et al. (Accepted) introduce a latent network model for double-sampled data, which simultaneously estimates a true underlying network of directed ties and error parameters for each reporter, and directly incorporates mutuality. Our contribution involves an improved model for a latent network that accommodates *any number of reporters*, allows directed ties, and incorporates mutuality explicitly into the generative model of reports.

Our proposed model also requires a new estimation algorithm, which is an additional contribution of our work. This is because previous generative models for multiply reported data can be written as a finite mixture (Titterington et al., 1985) of probability distributions. For example, the probability distribution for a present tie could be different than the probability distribution for an absent tie. Finite mixture models can often be estimated with efficient algorithms, such as expectation-maximization, and have been used in network research where data come from unreliable reporters (Butts, 2003) or feature a significant amount of missingness (Peixoto, 2018). The unique formulation of our model requires an infinite mixture model approach and standard methods cannot be easily applied. Therefore, we propose a generative model for latent networks that simultaneously handles multiply reported ties and weighted reports, while allowing individuals to vary in reliability. To estimate our model, we introduce an efficient variational inference algorithm.

2 The model

Consider the problem of collecting a network of ties between individuals. These ties could, for instance, represent relationships commonly studied in the social sciences—such as loaning money, giving advice, or sharing food. This can be done by querying a set of M reporters about the existence of ties. The real network is not observed; responses of the reporters are the only observed data at our disposal. We assume that the unobserved network is correlated with these responses. Mathematically, we define this as an $N \times N$-dimensional adjacency matrix, Y, where entries $Y_{ij} \in \{0, 1, \dots\}$ indicate the weight of the tie $i \to j$. For each tie type, the observed data is an $N \times N \times M$-dimensional tensor, X, with entries X_{ijm} containing reports by respondent m about the tie $i \to j$.

We assume that each respondent can, in principle, report on any tie within the network. The exact rule of how reporters respond may change with the application, but may be flexibly represented by a binary mask, R, of entries R_{ijm}. We set $R_{ijm} = 1$ whenever a reporter, m, is surveyed about the possible existence of a tie from node i to node j, and set the entry to 0 otherwise. In scenarios where a network has been double-sampled—e.g., where the same reporter responds about *giving* and *receiving* social support—every tie type is sampled twice (for each reporter), once for each direction of the interaction. These binary masks are convenient in the inference procedure as they remove the contributions of nonreporters.

As an example, m can nominate who she gives advice to (giving) and who she receives advice from (receiving). In this case, $m \in \{i, j\}$, and we distinguish the direction of the reported data using the notation X_{ijm} to indicate i to j flows and X_{jim} to indicate j to i flows. While we gave an example for ties of type *advice*, the model applies for any type of directed tie. To keep the model flexible, we model weighted ties with positive and discrete weights, so that $X_{ijm} \in \{0, 1, \dots\}$. This also includes the binary case, when X_{ijm} captures only whether a tie exists or not.

One of the main objectives of our model is to estimate the structure of a latent network, Y, from the reported data, X. Note that the term 'latent network model' is also used for models predicting network ties that incorporate latent variables to account for tie dependence implicitly (e.g., latent space models; Hoff et al., 2002). In contrast, we are modelling networks whose ties are unobserved or latent. We adopt a probabilistic approach where we assume that X depends on Y in a potentially noisy way. This means that we infer a probability distribution over possible generative structures compatible with the reported ties. We assume *conditional independence* between the entries of X, *given* Y, and the model's parameters. This is a common assumption made in network models (e.g., Newman, 2018b; Peixoto, 2018; Young et al., 2021), and makes estimation of the model more tractable. Typical exceptions where this assumption may not hold are scenarios where an upper limit is set on the maximum number of nominations a reporter can make—e.g., when respondents are asked: *Who are your five closest friends?*. In these scenarios, there is a (weak) negative correlation between nominations, because the likelihood of future nominations is reduced each time a nomination is made by a respondent, simply because the respondent is strictly limited to an arbitrary, finite set of nominations (Hoff et al., 2013). While this is important to note, solving this problem is beyond the scope of the current manuscript.

A further core objective for our model is to estimate the reliability of reporters. Reporters may under-report (i.e., neglect to report a tie, when it does exist) or over-report (i.e., report a tie, when it does not exist), and we account for these biased reports by assigning a 'reliability' parameter, θ_m, to each reporter m. For ease of interpretability, we think of this parameter as a positive number taking higher values when the reporter exaggerates their reports and lower values when they under-report.

Finally, we incorporate the intuition that reporters tend to nominate the same people for both directions of a relationship, X_{ijm} and X_{jim}. We term this pattern 'mutuality', to keep the concept distinct from the standard concept of dyadic reciprocity (henceforth termed reciprocity) in the true unobserved network Y. Bringing all of these modelling consideration together, we posit that the expected value of the data can be given as

$$\mathbb{E}\left[X_{ijm} | Y_{ij} = k\right] = \theta_m \lambda_k + \eta X_{jim}, \tag{1}$$

where $\eta \geq 0$ is the mutuality parameter. Mutuality enters the model as an additive and positive contribution to the expected number of reported ties. This measures the possible increasing weight of a directed tie, given that we observe the same tie in the opposite direction, as reported by the same reporter. The parameter λ_k is a positive real value that needs to be inferred, which regulates the contribution of Y in determining X. Note that the index k here refers to the positive and discrete value posited for Y_{ij}. In case of binary entries, $k \in \{0, 1\}$, but in this work, we assume more generally $k \in \{0, 1, \ldots\}$.

From this, we note how, for a given value of $\lambda_k > 0$, reporters with high θ_m tend to nominate more individuals, while reporters with smaller values tend to nominate fewer individuals. In contrast, a $\theta_m = 1$ indicates a neutral contribution (neither over-reporting nor under-reporting), hence we can interpret it as representing an unbiased reporter. Regardless of the reporter's 'reliability', the existence of a tie X_{jim} in one direction increases the expected value of X_{ijm} in the opposite direction, when $\eta > 0$. This also implies that it may not be possible to identify the reliability of reporters that report a high percentage of ties in both directions, and in networks with high values of η. In these cases, in fact, the presence of a reported tie can be determined with a high likelihood based on the tie reported in the opposite direction.

To form a likelihood for the observed data that can accommodate various network and report structures—in particular, directed and weighted networks—we write the conditional distribution:

$$P(X_{ijm} \,|\, X_{jim}, Y_{ij} = k, \lambda_k, \theta_m, \eta) = \frac{(\theta_m \lambda_k + \eta X_{jim})^{X_{ijm}}}{X_{ijm}!} \mathrm{e}^{-(\theta_m \lambda_k + \eta X_{jim})}. \tag{2}$$

Note that this choice of a Poisson distribution leads to an expected value for X_{ijm} as in equation (1). Furthermore, the positivity of the parameters makes this expression valid without the need of a link function. From this conditional, one can specify a two-point joint likelihood of (X_{ijm}, X_{jim}) by suitably defining the marginal distribution $P(X_{jim} \,|\, Y_{ji} = k, \lambda_k, \theta_m, \eta)$. While there exist choices resulting in a consistent joint likelihood (see Section S1.3 for details), these may not result in simple, efficient closed-form updates of the parameters. Hence, we assume a pseudo-likelihood approximation (Besag, 1974) for the two-point likelihood, as done in Safdari et al. (2021)

$$\begin{aligned} &P(X_{ijm}, X_{jim} \,|\, Y_{ij} = k, Y_{ji} = q, \lambda_k, \lambda_q, \theta_m, \eta) \\ &\quad \approx P(X_{ijm} \,|\, X_{jim}, Y_{ij} = k, \lambda_k, \theta_m, \eta) \times P(X_{jim} \,|\, X_{ijm}, Y_{ji} = q, \lambda_q, \theta_m, \eta). \end{aligned} \tag{3}$$

The model can be applied to any tie type encoded in the input data X, and it will output the reliability of a reporter for that tie type. One can potentially generalize this to a multi-layer framework by considering a unique θ_m for each reporter, regardless of tie type. This would then introduce a coupling between the reported X for various tie types, potentially increasing the complexity of the model. Alternatively, one could consider a different θ_m for each tie type. If these different types of reliability are considered independent from each other, then our model could be readily generalized to include these distinctions, without need for further extra coupling, but only additional distinct priors. This is essentially equivalent to running our model on each layer (i.e., tie type) individually, as we do in our numerical experiments on real data below.

Potentially, one could also include a different θ_m depending on the directionality of the ties—i.e., a $\theta_m^{\rightarrow}$ for ties sent and a $\theta_m^{\leftarrow}$ for ties received—capturing situations where reporters could over-report in one direction and under-report in another one. This would modify equation (3) to contain one of these two parameters inside the corresponding conditional distribution. If $\theta_m^{\rightarrow}$ and $\theta_m^{\leftarrow}$ are thought to be independent, so that their priors factorize, then this would lead to a straightforward generalization of the algorithm.

We assume that there are no contributions to the likelihood of X when a reporter is censored—i.e., when m is not given the chance to report on the tie $i \rightarrow j$. In empirical applications, this could arise, for example, when a survey design only asks about ties directly involving the reporter.

In addition to specifying the likelihood as in equation (2), we adopt a Bayesian approach and assume priors for the parameters and the unobserved Y. To maximize the flexibility of our model, we allow for positive and discrete values of Y by using a categorical prior

$$P(Y_{ij} = k; p_{ij}) = p_{ij,k}, \tag{4}$$

where p_{ij} is the parameter of the categorical prior distribution, and $\sum_k p_{ij,k} = 1$. The sum runs over the possible positive and discrete values of Y_{ij}. The resulting model can thus accommodate, for example, a binary network Y and weighted reports X, as the likelihood in equation (3) is valid for any number of values that Y can take. We then consider Gamma priors for the remaining parameters, as they are defined for positive real numbers, and are conjugate with the Poisson distribution, which makes calculations convenient.

3 Inference

Because of the possibility of mutuality in nominations, we do not have a closed-form joint distribution for (X_{ijm}, X_{jim}), hence we consider the conditional distribution

$$\begin{aligned} P\big(\{X_{ijm}\}_m | \{X_{jim}\}_m, Y_{ij}, \lambda, \{\theta_m\}_m, \eta\big) &= \prod_m P(X_{ijm} \mid X_{jim}, Y_{ij}, \lambda, \theta_m, \eta) \\ &= \prod_k \left[P(Y_{ij} = k) \prod_m P(X_{ijm} \mid X_{jim}, Y_{ij} = k, \lambda_k, \theta_m, \eta) \right]^{Y_{ij,k}}. \end{aligned} \tag{5}$$

By using a pseudo-likelihood approximation as in Safdari et al. (2021), the full posterior can be written as

$$\begin{aligned} P(Y, \lambda, \theta, \eta \mid X) \propto P(X \mid Y, \lambda, \theta, \eta) P(Y) P(\lambda) P(\theta) P(\eta) \\ = \prod_{i,j} P\big(\{X_{ijm}\}_m | \{X_{jim}\}_m, Y_{ij}, \lambda, \{\theta_m\}_m, \eta\big) P(Y_{ij}; p_{ij}) \\ \prod_k P(\lambda_k; a_k, b_k) \prod_m P(\theta_m; \alpha_m, \beta_m) P(\eta; c, d) \end{aligned} \tag{6}$$

$$=: \mathcal{L}(\lambda, \theta, \eta, Y) \tag{7}$$

where the proportionality results from the omission of an intractable normalization that does not depend on the parameters. To estimate the model, we use variational inference with a mean-field variational family (Blei et al., 2017), which yields an approximate posterior distribution for the network and parameters. The algorithmic updates needed to find the best approximation to the posterior distribution follow a coordinate ascent routine, iteratively finding the best marginal posterior distribution of each parameter while holding the others fixed. We call the resulting algorithm VIMuRe, for Variational Inference for Multiply Reported data. The model is efficient, as it exploits the sparsity of the dataset. Specifically, the numerical implementation has a computational complexity that scales linearly with the number of nonzero entries of R, the reporters' mask, typically a sparse quantity. As a comparison, techniques based on sampling (e.g., HMC) can take an order of magnitude longer to run (see Blei et al., 2017), depending on the underlying complexity of the model. As the output results depend on the random initial configuration of the parameters, we run the algorithm several times and then consider the realization that resulted in the best ELBO value, as usually done in variational inference. This makes the output robust against initial values, as we expect a decreasing sensitivity to them for increasing $N_{\text{realisations}}$. In our experiments, we found that already $N_{\text{realisations}} = 5$ was a reasonable value to guarantee robust results. Pseudo-code for the algorithm is shown in Algorithm 1; see Section S1.1 for further details.

4 Simulation experiments

To validate our model, and study its performance in different regimes, we simulate synthetic data that reproduce our analysis scenarios—multiply reported network data that depend on a latent adjacency matrix—using the model itself. In detail, we first generate the network Y either with a flexible version of a mixed-membership stochastic block model (MULTITENSOR, De Bacco et al., 2017), a degree-corrected stochastic block model (DC-SBM, Karrer & Newman, 2011), or a probabilistic model with reciprocity (CRep, Safdari et al., 2021). We then generate the observed X given fixed reliability, mutuality, the generated network and the contribution of λ, collectively denoted by $\Theta = (Y, \theta, \lambda, \eta)$. We follow the approach described in Safdari et al. (2021), and for each reporter m we draw a pair (X_{ijm}, X_{jim}) consistently with the joint $P(X_{ijm}, X_{jim}|\Theta)$ in a two-step sampling routine, where we first generate one of the two reported ties and then the second one given the first, see Sections S1.2 and S1.3 for details.

In the simulations, we examine our ability to recover: (i) the underlying network, Y, and (ii) the individual reliabilities, θ_m. First, we generate synthetic networks reproducing three different scenarios. Two of these scenarios are extreme cases, where a fraction of reporters (θ_{ratio}) are tagged to be either over-reporters, or under-reporters, while all the others are reliable—i.e., they have $\theta_m = 1$, and their X entries are deterministically generated. In doing this, we document model performance in difficult cases, where the proportion of unreliable reporters is high. The third scenario, is more realistic. In this setting, we have both over- and under-reporters, as we draw θ_m from a Gamma distribution, providing a broad range of values. We vary the difference between λ_1 and λ_0, such that the smaller this difference becomes, the noisier the problem gets, and thus the harder the inference tasks. Secondly, we investigate the ability of our model in recovering structural properties of the latent network Y—e.g., reciprocity, density, and communities—in other sets of synthetic networks.

In all experiments, we fit two versions of the model: a version with mutuality (VIMuRe$_T$) and a version without (VIMuRe$_F$). To provide a point of comparison, we also compute two baselines estimates of Y: (i) the *union*, in which a tie exists if at least one reporter reports that tie, and (ii) the *intersection*, in which all the reporters of a tie have to agree for the tie to exist. These two commonly used baselines represent the most and least inclusive approaches to integrating multiply reported data, and so provide reasonable comparisons for VIMuRe.

4.1 Results

We use the F1-score—i.e., the harmonic mean of precision (fraction of inferred ties that actually exist) and recall (fraction of existing ties found by the method) measures—to assess the ability of our model to recover Y, which is binary in our experiments. This choice is chiefly motivated by the fact that we have unbalanced data (since many fewer ties than possible tend to exist in empirical networks), that the F1 score is widely understood, and that our inferences based on F1 scores are qualitatively identical to those based on the Matthews correlation coefficient (Chicco & Jurman, 2020).

Algorithm 1: VIMuRe.

Input: Data X, Model $\mathcal{L}$, Variational family q.
Initialize the variational parameters $\gamma,\ \phi,\ \rho,\ \nu$ to the priors with a small random offset.
while *change in ELBO is above a threshold* **do**

For end each pair of nodes such that $X_{ijm} > 0$, update the multinomials:

$$\hat{z}^1_{mk} \propto \exp\left\{\Psi(\gamma^{\text{shape}}_m) - \log\gamma^{\text{rate}}_m + \Psi(\phi^{\text{shape}}_k) - \log\phi^{\text{rate}}_k\right\}$$

$$\hat{z}^2_{ijm} \propto \exp\left\{\Psi(\nu^{\text{shape}}) - \log\nu^{\text{rate}} + \log X_{jim}\right\} = X_{jim}\exp\left\{\Psi(\nu^{\text{shape}}) - \log\nu^{\text{rate}}\right\}$$

where the proportionality is such that $\hat{z}^1_{mk} + \hat{z}^2_{ijm} = 1$.
For each reporter, update the reliability parameters:

$$\gamma^{\text{shape}}_m = \alpha_m + \sum_{i,j,k} R_{ijm}\,\rho_{ij,k}\,X_{ijm}\,\hat{z}^1_{mk}$$

$$\gamma^{\text{rate}}_m = \beta_m + \sum_{i,j,k} R_{ijm}\,\rho_{ij,k}\,\frac{\phi^{\text{shape}}_k}{\phi^{\text{rate}}_k}\ .$$

For each possible value k of Y_{ij}, update the parameters:

$$\phi^{\text{shape}}_k = a_k + \sum_{i,j,m} R_{ijm}\,\rho_{ij,k}\,X_{ijm}\,\hat{z}^1_{mk}$$

$$\phi^{\text{rate}}_k = b_k + \sum_{i,j,m} R_{ijm}\,\rho_{ij,k}\,\frac{\gamma^{\text{shape}}_m}{\gamma^{\text{rate}}_m}$$

and:

$$\rho_{ij,k} \propto \exp\left\{\log p_{ij,k} + \sum_m R_{ijm}\left(X_{ijm}\,\hat{z}^1_{mk}\,\mathbb{E}_{q(\lambda_k)}[\log\lambda_k]\right) - \frac{\phi^{\text{shape}}_k}{\phi^{\text{rate}}_k}\sum_m R_{ijm}\,\frac{\gamma^{\text{shape}}_m}{\gamma^{\text{rate}}_m}\right\}\ .$$

Update the mutuality parameters:

$$\nu^{\text{shape}} = c + \sum_{i,j,k,} \rho_{ij,k}\sum_m R_{ijm}\,X_{ijm}\hat{z}^2_{ijm}$$

$$\nu^{\text{rate}} = d + \sum_{i,j,m} R_{ijm}\,X_{jim}.$$

end
Output: Variational parameters $(\gamma,\ \phi,\ \rho,\ \nu)$.

For readers more familiar with the latter, we include Matthews correlation coefficient results in Supplementary Materials.

In the two extreme scenarios, where there are only over- or under-reporters, our model recovers the unobserved network, Y, better than approaches that take the union or intersection of the reported ties in X. The performance of our model is also more robust as the number of unreliable reporters and/or mutuality increases, see Figure 1. In particular, our model with mutuality (VIMuRe$_T$) has a higher performance for high values of η, which is also a harder regime, as the performance of all methods decreases in this range. In general, the performance of the baselines decrease as the number of over- or under-reporters grows. For example, the union baseline estimates relationships that do not exist in the true network, when there are several over-reporters. Conversely, the intersection baseline underestimates the amount of ties, when a high fraction of individuals under-report. Our model overcomes these biases by accounting for reporters'

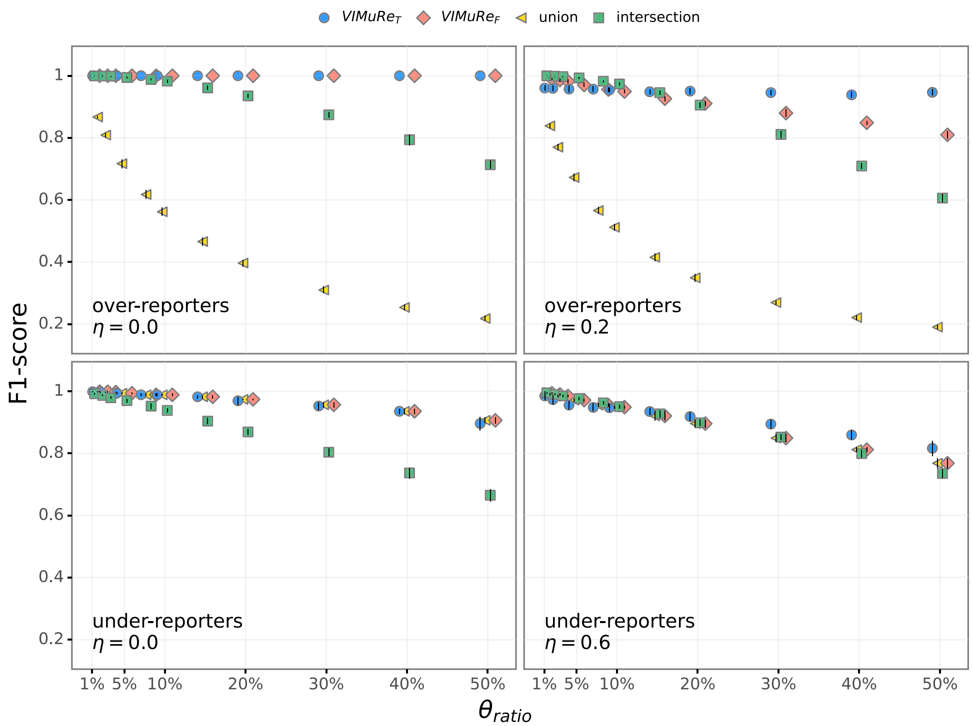

Figure 1. Estimating underlying network, Y, in synthetic networks with over- or under-reporters. Synthetic networks with $N = 100$ nodes and $M = 100$ reporters, generated with the benchmark generative model described in Sections S1.3 and 4, by varying the fraction θ_{ratio} of over-reporters (top) or under-reporters (bottom). The two columns represent networks generated without (left) and with (right) the mutuality effect η. The results are averages, and standard deviations calculated over ten independent synthetic networks. The accuracy of the estimate of the underlying network, Y, is measured with the F1-score. This measure ranges from 0 to 1, where 1 indicates perfect matching. See Figure S1 for similar plots based on the Matthews correlation coefficient and Figures S4 and S5 for additional experiments where M varies.

reliability, and this results in higher and more robust performance. However, when θ_{ratio} becomes too large, VIMuRe also fails since, as Figure S2 shows, recovering the reporters' reliability becomes harder. That said, the model with mutuality performs better at this task, and fails much more slowly than the model without mutuality, especially when η is large. To assess robustness in recovering Y as the number of reporters varies, we run further experiments keeping the same settings as above for $N = 300$ and varying $M \in [25, 300]$. For this simulation, we fixed $\theta_{\text{ratio}} = 0.50$, thus capturing the most challenging case explored in the original simulations in Figure 1. We find that while performance decreases as the number of reporters decreases, as expected, VIMuRe$_T$ captures the ground truth of Y better than baseline implementations across different M, as shown in Figure S4.

Performance differences are more nuanced when we consider the more realistic experiment, which features a broad range of reporter reliabilities. F1-scores are lower than in the previous experiments, in general, and recovering the ground truth is particularly challenging when the difference between the mean number of reports of a tie being present and not, $\lambda_1 - \lambda_0$, is lower, see Figure 2. Intuitively, as the difference $\lambda_1 - \lambda_0$ decreases, both the zero and nonzero inputs of Y tend to make the same contribution in determining X; thus, it becomes more difficult to distinguish true ties on the basis of reports. These experiments also further confirm what we observed in the previous experiments, that the hardest regime features the highest mutuality. A higher η means that a reporter will tend to nominate the same set of people for both *giving* and *receiving* questions, which results in X having less informative information. In these experiments, both versions of our model and the union baseline perform similarly while the intersection baseline performs much

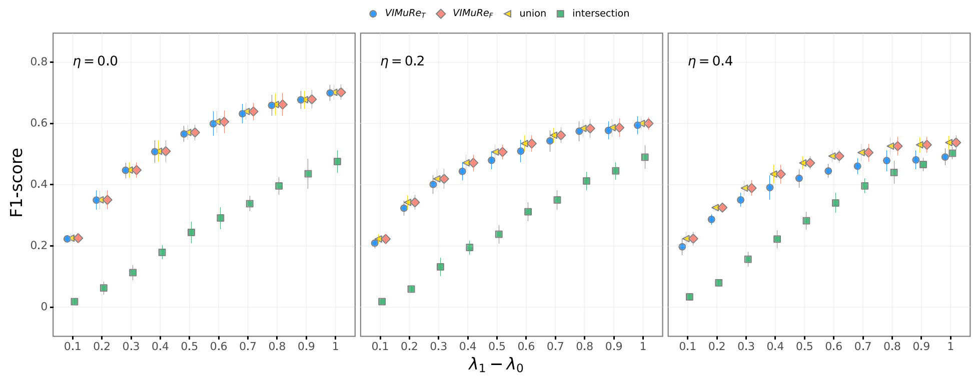

Figure 2. Estimating underlying network, Y, in synthetic networks with over- and under-reporters. Synthetic networks with $N = 100$ nodes and $M = 100$ reporters, generated with the benchmark generative model described in Sections S1.3 and 4, by varying the difference between λ_0 and λ_1. The three columns represent networks generated with no (left), medium (centre), and high (right) mutuality, η. The results are averages and the standard deviations over ten independent synthetic networks. The accuracy of the estimate of the underlying network, Y, is measured with the F1-score. This measure ranges from 0 to 1, where 1 indicates perfect matching. See Figure S3 for similar plots based on the Matthews correlation coefficient.

more poorly. However, the performance gap decreases as mutuality increases and more ties are reported. Further studies on the recoverability of the latent network mutuality in synthetic experiments are provided in Section S1.4 and Figure S6.

Once we have an estimate, $\widehat{Y}$, of the unobserved network, Y, a practitioner would be able to investigate structural properties in the latent network. To give an example, we will assess the ability of our model to capture reciprocity on the estimated $\widehat{Y}$, a foundational feature of many social relations (Fehr & Gächter, 1998; Molm, 2010). To this end, we convert the posterior probability distribution $\rho_{ij,k}$ to a binary unweighted adjacency matrix. Since we considered binary data in our experiments, and thus $k \in \{0, 1\}$, we can obtain this by applying a threshold to the sub-tensor $\rho_{ij,k=1}$, as it represents the probability distribution of finding a $\widehat{Y}_{ij} = 1$ entry. After each run of VIMuRe$_T$, we apply a range of thresholds $t_\rho \in [0.050, 0.075, \ldots, 0.725, 0.750]$ such that we assign $\widehat{Y}_{ij} = 1$ when $\rho_{ij,k=1} \geq t_p$, and keep track of the best t_ρ^*, for which reciprocity in the inferred network most closely matches the reciprocity of the ground truth. In Figure 3, we show that the relationship between this optimal threshold and the mutuality, η_{est}, as inferred by VIMuRe$_T$, can be approximated by the linear equation

$$t_\rho^* = 0.33\eta_{\text{est}} + 0.10. \tag{8}$$

In fact, in Figure 4, we show that VIMuRe$_T$ outperforms all other models at this task when the threshold on ρ was set according to the heuristic proposed above. Reciprocity estimated by the model was a closer match to the reciprocity of the true unobserved network even in simulations with high values of mutuality, a scenario where other methods tend to overestimate reciprocity. Density of the inferred network is also closer to ground truth, when compared with baseline methods in most scenarios, as can be seen in Figure S7. These results mean that despite the small gap on the value of F1-score, VIMuRe$_T$ may be able to provide a good estimate of structural properties of Y. The values of t_ρ^* in equation (8) are valid for the settings considered in the experiments analysed here and reported in Table S1, which control how synthetic networks are generated. Being a heuristic, one can in principle obtain a different formula when simulating data under different assumptions (e.g., varying N, M or reciprocity).

As a final test, we also show that VIMuRe allows the underlying community structure of a network to be recovered, even when it is measured noisily. To this end, we use a latent network that has planted overlapping communities, and generate reports as before. We then estimate the network, and finally recover communities using a probabilistic generative model with latent variables (De Bacco et al., 2017). Figure 5 shows the result of this experiment, and illustrates that VIMuRe is

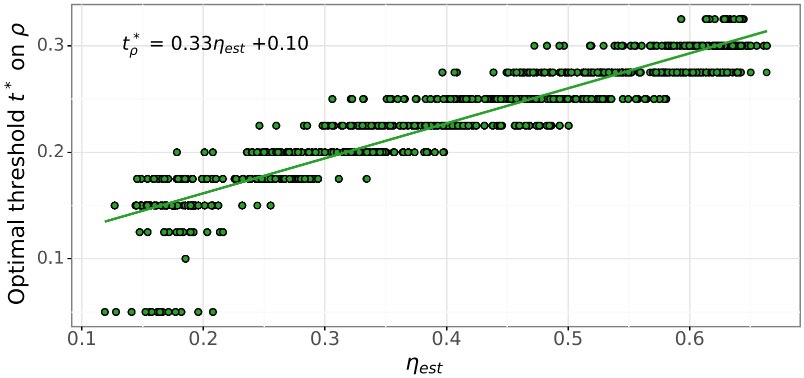

Figure 3. Synthetic networks with $N = 100$ nodes and $M = 100$ reporters, generated with the benchmark generative model described in Sections S1.3 and 4 with $\lambda_1 - \lambda_0 = 1.0$, and planted reciprocity values around ≈ 0.2 on the ground truth network, Y. The plot shows that the threshold that best captures reciprocity is linearly correlated with η_{est}.

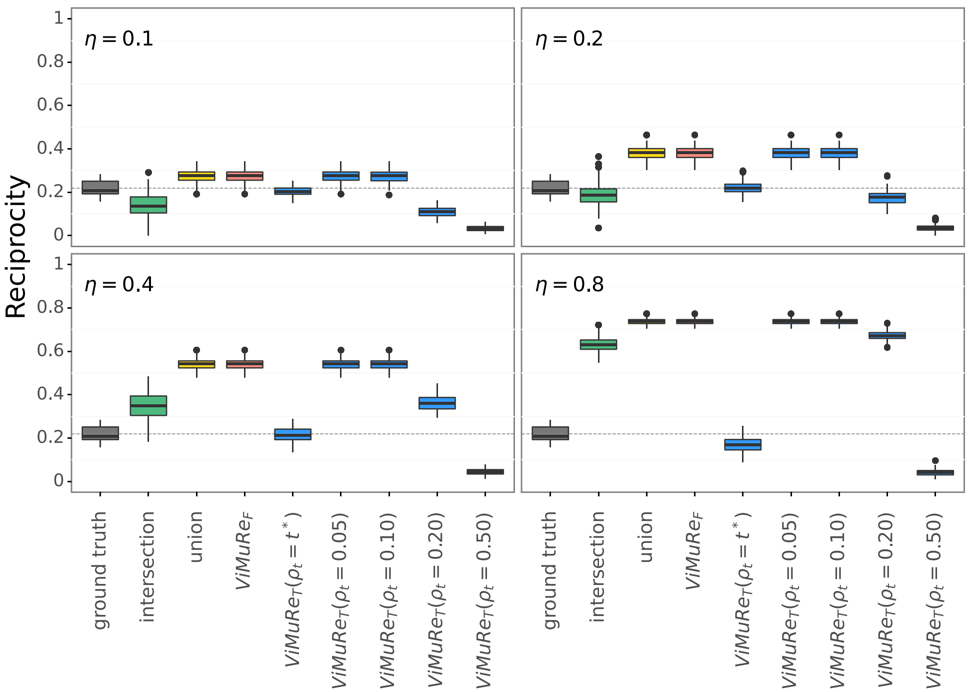

Figure 4. Reciprocity recovery from synthetic networks with $N = 100$ nodes and $M = 100$ reporters, generated with the benchmark generative model described in Sections S1.3 and 4 with $\lambda_1 - \lambda_0 = 1.0$, and planted reciprocity values around ≈ 0.2 (horizontal dashed line) on the ground truth network, Y. The four sub-plots represent networks generated with low (top left), medium (top right), to increasingly high (bottom left and right) mutuality effects, η. The box plots are distributions of the reciprocity in Y, over a sample of one hundred synthetic networks.

more robust than other approaches across different mutuality values, providing slightly better results than all other models; the intersection performs the worst. The qualitative example on the right panel of Figure 5 highlights how VIMuRe$_T$ infers a partition closer to the ground truth than those inferred by the other methods, especially when mutuality is higher.

To summarize, our simulation experiments suggest that the use of a generative model with latent variables results in more robust estimates of the true underlying network, Y, in comparison to

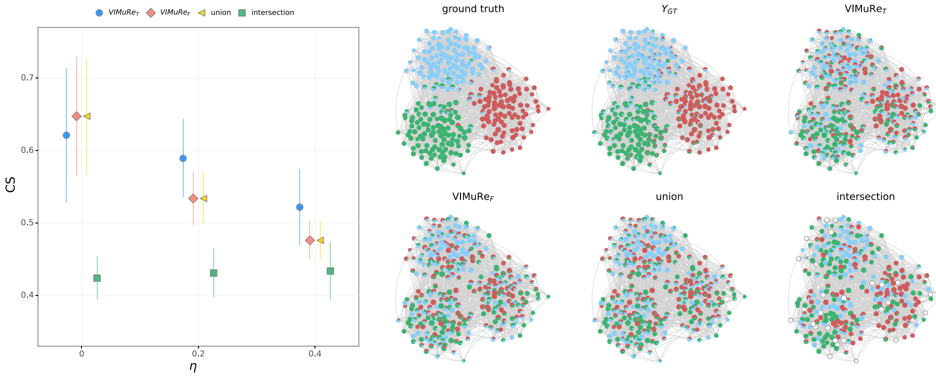

Figure 5. Community structure recovery from synthetic networks with both over- and under-reporters. Synthetic networks with $N = 300$ nodes and $M = 300$ reporters, generated with the benchmark generative model described in Sections S1.3 and 4 with $\lambda_1 - \lambda_0 = 1.0$. The results shown in the left frame are averages and standard deviations over ten samples of synthetic networks, generated by varying the mutuality parameter, η. The accuracy in recovering the overlapping community structure is measured with the cosine similarity (CS) using the inferred membership vectors. In the right frame, we plot examples of the partitioning of a synthetic network, generated with $\eta = 0.2$. The 'ground truth' is the partition used to generate Y, and Y_{GT} stands for the partition found by using the true Y. Nodes colored in white represent isolated nodes.

deterministic approaches (such as taking the union or intersection of sub-tensors). Furthermore, our model yields an estimate of reporter reliabilities, which can provide additional insights about the data-generating process. In addition, we note that our model performs better than other models when we include the mutuality parameter, η. In particular, VIMuRe$_T$ shows better results in estimating reciprocity than VIMuRe, specifically in cases where people's propensity to report mutuality in their relationships is high.

5 Analysis of Nicaragua data

We apply our modelling approach to data collected from a horticulturalist community in Nicaragua (see Koster, 2018, for more detail on the population and measurement instruments). These data were collected using a roster-based design, where all adult residents within the community were presented with a list of all other adult residents, and were asked two questions about relationships related to social support (i.e., *Who provides tangible support to you at least once per month?* and *Who do you provide tangible support at least once per month?*). Previous studies have performed separate analyses on the two questions (Koster, 2018; Simpson, 2022). We examine both questions in a single model, examining the potential biases that shape the reports of social support.

In this dataset, the reports vary significantly across reporters, with some reporters nominating many ties and others nominating fewer. It is, therefore, reasonable to have the priors on θ_m reflect this. We can incorporate this insight by running the inference in two steps, where we first run VIMuRe with a weak prior that is the same for all reporters, while in a second step we run VIMuRe with a prior proportional to the posterior mean of θ_m inferred in the first step. This is in line with Empirical Bayes approaches (Casella, 1985; Morris, 1983; Robbins, 1955) that estimate prior distributions from the data. This approach allows us to obtain a wider range of reliabilities so that we can better distinguish possible exaggerators than when using the same prior for all reporters.

Applying VIMuRe produces estimates of a network—which is binary and obtained by applying the optimal threshold in equation (8)—that has properties (e.g., mean degree, reciprocity) that fall somewhere between the results of taking the union (which returns an incredibly dense network) and taking the intersection of the double-sampled ties (which returns an extremely sparse network). See Table S2, for a summary. Overall, mutuality was estimated to be $\eta_{\text{est}} = 0.540$ and reciprocity was 0.11.

In contrast to other survey data such as name generators, where survey design may contribute to under-reporting, the roster-based design makes it much easier for respondents to make many nominations. In the roster design in Nicaragua, informants reported approximately 25 alters for each prompt, substantially more than the average of 4 from the constrained name generators used in the

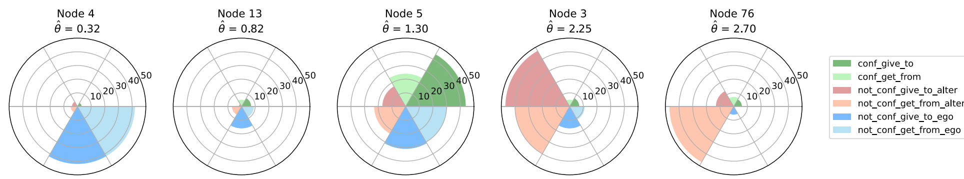

Figure 6. Example of individual reliabilities. Pie plots show six different configurations for the reported ties (two per each direction of a tie): ties confirmed by both reporters (conf 'give to', conf 'get from'); ties reported by m but not confirmed by others (not conf 'give to' (alter), not conf 'get from' (alter)); ties reported by others but not by m (not conf 'give to' (ego), not conf 'get from' (ego)). Each plot is a different reporter; their estimated reliability $\widehat{\theta}$ is printed on top. Each slice of the pie is one tie reported in one the six possible ways, represented by the colours. In this example, we consider reporters from the Nicaragua dataset.

study described below in Section 6 (see Table S2 for full network statistics). As can be seen in Figure 6, however, we nevertheless observe reporters with low θ_m, who were nominated by several others, but nominated relatively few themselves (e.g., Nodes 4 and 13). On the opposite extreme, we see reporters with high θ_m, who nominated many others, but whose ties are not confirmed by those alters (e.g., Nodes 3 and 76). In between, we show an example of a reporter (Node 5) with intermediate value of θ_m, who nominates several others in a way consistent with the reports of others. The distribution of reliability for reporters in this dataset can be seen in Figure S8.

Since these data are not explicitly generated with the generative model assumed by VIMuRe, we run a goodness-of-fit test to ensure that the model is appropriate for the above analysis. To do this, we use a series of posterior-predictive checks (Gelman et al., 1996, 2013), which compare the Nicaragua data with synthetic data $\widetilde{X}$ generated using the fitted model. The posterior-predictive distribution is defined as

$$P(\widetilde{X} \mid X) = \sum_{Y} \iiint P(\widetilde{X} \mid Y, \lambda, \theta, \eta) P(\lambda, \theta, \eta, Y \mid X)\, d\lambda\, d\theta\, d\eta \tag{9}$$

and one can generate samples from this distribution by first obtaining samples $(Y, \lambda, \theta, \eta)$ from the variational approximation to the posterior distribution, and then using these parameters as input to create new synthetic data, $\tilde{X}$, from the likelihood described in Section 2. A good fitted model should lead to new synthetic data $\tilde{X}$ that resembles the input X. We run two numerical posterior-predictive tests to assess the appropriateness of the VIMuRe model: (i) a direct comparison between the elements of the Nicaragua data X and the distribution of $\tilde{X}$, and (ii) a test checking whether two samples from the posterior-predictive distribution are typically more, equally, or less distant from one another than a sample from the posterior-predictive distribution and the Nicaragua data (Young et al., 2021). The results shown in Figure 7 confirm that the VIMuRe model is appropriate for our analysis.

6 Analysis of social support networks in Karnataka

To further highlight the broad applicability of our modelling approach across elicitation methods, we apply VIMuRe to a dataset of social support networks collected from 75 villages in the Indian state of Karnataka (Banerjee et al., 2013). As part of a larger project, a series of name generators were asked of most of the adult members of a subset of households in each village (overall, about 46% of all households were surveyed). The name generators included questions about four double-sampled relationships: who people give advice to or receive advice from (*Advice*), who people would borrow from or lend a small amount of money to (*Money*), who people go to or receive as visitors (*Visit*), and who people would borrow kerosene and rice from or lend kerosene and rice to (*Household Items – 'HH Items' in the plots*). In the past, these data have been studied by aggregating responses from multiple household respondents and taking the union of the double-sampled questions (Banerjee et al., 2013; Jackson et al., 2012).

Our results suggest that the reciprocity values in these networks are in fact lower than what would be obtained by simpler approaches, as shown in Figure 8, generally, and illustrated for one specific village and tie type in Figure 9. While we do not have ground truth values in this

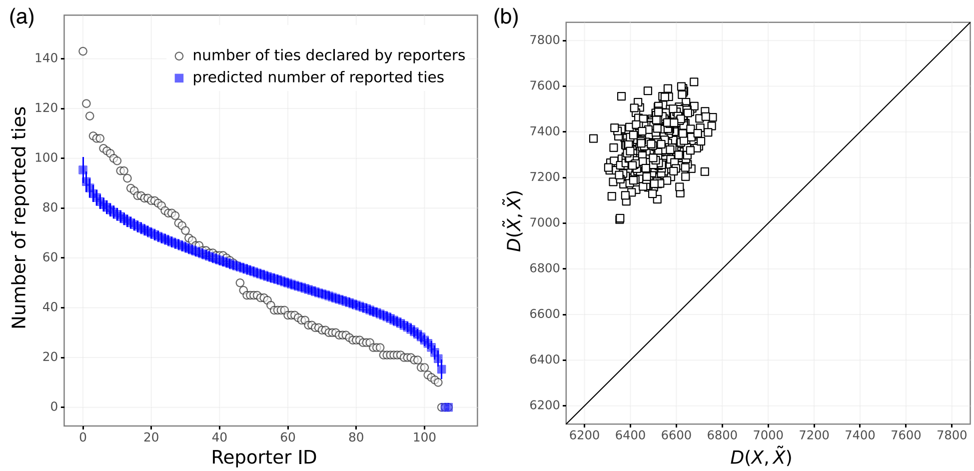

Figure 7. Example goodness-of-fit analysis for the Nicaragua dataset. (a) Number of ties declared by each reporter (squares) across the double-sampled social support question, compared with the average predicted number of reported ties (circles). Error bars correspond to standard deviations computed with $n = 500$ samples from the posterior distribution. (b) Scatter plot showing a model-model ($D(\tilde{X}, \tilde{X})$) versus model-data ($D(X, \tilde{X})$) comparison. Each dot corresponds to one posterior-predictive sample and illustrates the distance between this sample and the Nicaragua data on the horizontal axis (model-data), and another random posterior-predictive sample on the vertical axis (model-model). The model can be deemed appropriate when these two distances are similar—i.e., if the scatter plot is a point-cloud centred on or close to the diagonal (Young et al., 2021). We selected the Hamming distance D as the test statistics, defined as the number of pairs of entries (X_{ijm}, $\tilde{X}_{ijm}$) in disagreement between two datasets.

case, we note that these numbers are similar to those obtained on the synthetic networks in our experiments shown in Figure 4. In particular, they mimic the situation with high mutuality, where the union and intersection significantly overestimate the reciprocity on Y, whereas VIMuRe identifies the correct range of values. These results suggest that reciprocity will likely be overestimated in double-sampled network data, when the reports have high mutuality. Of the four tie types in the data from Karnataka, the estimations made by VIMuRe suggest that the 'Advice' layer has the lowest reciprocity values on average (0.39 ± 0.07), with 'Money' (0.43 ± 0.08), 'Visit' (0.47 ± 0.07), and 'Household items' (0.48 ± 0.08) layers exhibiting higher reciprocity.

Note, too, that our estimates for mutuality (η_{est}) follow a similar pattern, with the lowest estimates for 'Advice', and the highest for 'Visit' and 'Household Items'. These estimates broadly align with theory: reciprocity—and the *expectation* for reciprocity, as represented by the mutuality term —is higher in those relationships that are understood to be more balanced and mutually supportive (i.e., visiting one another's homes, and borrowing/lending basic household items, like rice and kerosene), and lower in those relationships that are potentially seen as hierarchical and imbalanced (receiving/giving advice, and borrowing/lending money).

We next investigate how reporter 'reliabilities' are distributed in these networks. Since the mutuality in these graphs is high ($\eta_{\text{est}} \geq 0.4$), it is very common that reporters repeat the same names across the different name generators. Therefore, it is expected that individual 'reliability' terms will play a smaller role in determining the reported social network. Recalling equation (1), this means that the value of θ_m will be small for reporters with a high rate of repeat nominations—that is, the proportion of alters reported by an ego on the 'give to' question that gets repeated on the 'gets from' question. In the Karnataka dataset, 26% (Advice) to 52% (Household Items) of reporters have an individual rate of repeated nominations of 100%. In such cases, small values of θ_m should not be interpreted as indicating under-reporting, as a small θ_m in this case is just a signal of a high mutuality. The vast majority of reporters (99.49%) with a small θ_m ($\theta_m < 0.1$) in the Karnataka networks have an individual rate of repeat nominations of 100%, regardless of the tie type.

For all four tie types, we observe that reporters tend to under-report relationships, even after having accounted for those individuals who have low θ_m for the reasons discussed above (see Figure S9). This is consistent with prior literature (e.g., Butts, 2003) that suggests that reporters

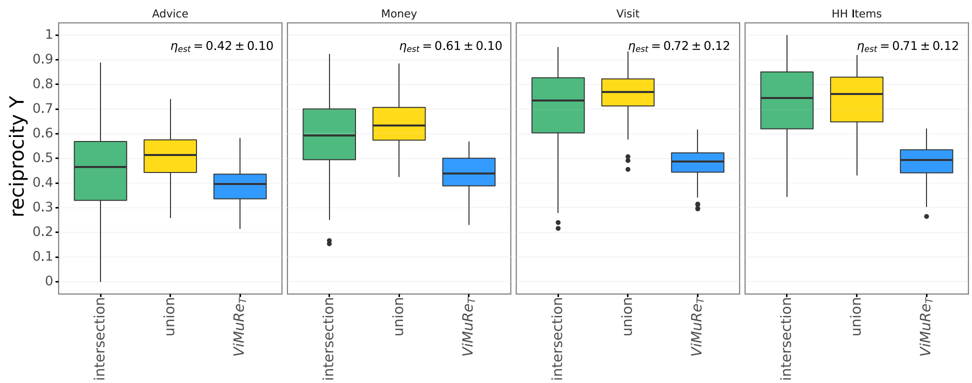

Figure 8. Reciprocity on Karnataka networks. The box plots show the distribution over the 75 networks of the reciprocity measured on the inferred $\widehat{Y}$. Each column is a different tie type, as written on the figure title.

Figure 9. Example of networks estimated by baseline methods and VIMuRe for one Karnataka village (tie type 'Visit'). (a) Union (recip. = 0.93), (b) intersection (recip. = 0.88), and (c) VIMuRe (recip. = 0.49).

are reasonably accurate when reporting ties, but quite inaccurate when reporting nonties. This effect may be partially induced by the way the questionnaire was formulated, as there were only four entries available for nominating alters. This causes under-reporting to be much more likely, and substantially limits the number of reported tie configurations that can be observed in this system. Indeed, we note a negative correlation between θ_m and the in-degree of ties reported by others involving m (see Figure S11). In particular, reporters nominated by many others (some by more than 20 people) could only nominate up to four among these; such reporters will necessarily have low values for θ_m.

We do not observe any strong differences in the distribution of reliability across the four tie types (see Figure S9). We assess whether reporters are consistent in their reliabilities across the tie types by examining the pairwise distances using the Wassertein distance (a metric for measuring distances between two distributions; Givens & Shortt, 1984) between each set of tie types (see Figure S10). We see some telling patterns by looking at the consistency of 'Advice' with the other tie types: while reporters are most consistent between the 'Advice' and 'Money' networks, they are least consistent between the 'Advice' and 'Household Items' or 'Visit' networks.

7 Discussion and conclusions

Self-report network data are an important resource for social scientists, but they are also susceptible to several types of reporter bias. Identifying the possible biases that structure self-reports of social relationships remains an essential and open area of research for social network analysis. Failure to identify and account for such reporting biases may lead researchers to draw incorrect inferences (Redhead et al., Accepted). But how should social scientists go about investigating and treating reporting bias in measurements of social networks? We provide a novel statistical solution to—and theoretical and empirical evidence of—this problem, with particular focus on two forms of bias. First, we investigate and adjust for the general propensity of reporting balanced, reciprocal relationships (i.e., mutuality). Second, we provide a method of accounting for individuals' unique potential to misrepresent or misreport their relationships during reports. Both of these forms of bias have the potential to add substantial 'noise' to empirical representations of social networks, but the extent to which this is present and problematic has not yet been well established. While previous work has explored individual propensities to misreport their ties (e.g., Butts, 2003; Newman, 2018a; Young et al., 2021), there has been limited formal analysis of the impact that 'mutuality' has on network inference (but see, Redhead et al., Accepted).

We have focused our attention on cases where multiple reporters are able to provide information on any given relationship. In particular, we have considered 'double-sampled' relationships, where respondents are asked about their role both as giver and as receiver—a common technique that is used in social support network surveys. We have introduced a probabilistic modelling framework, VIMuRe, that provides a principled solution to these issues and aims to more appropriately capture the data generating process associated with name generator designs. VIMuRe takes as input potentially biased, imperfect survey responses and uses these to estimate a 'true' latent network, as well as parameters governing individual biases and relationship-specific tendencies towards mutuality. The model estimates both a ground-truth, Y, and θ which, in certain cases, can be interpreted as a reporter's reliability (conditional on some level of mutuality).

The model that we have introduced here strongly departs from common approaches for dealing with double-sampled network data in the social sciences, in which researchers simply take the union or the intersection of nominations. Our approach also departs from existing network reconstruction methods and advances a framework that is maximally flexible. To our knowledge, existing network reconstruction methods (e.g., Butts, 2003; Newman, 2018a; Young et al., 2021) that are applicable to social networks focus on the single-sampled case—with the exception of Redhead et al. (Accepted), which is applicable only to double-sampled networks. While we have highlighted double-sampled network data here, our framework can be readily used for many reporting sampling schemes. A tie within a network could be reported on by any number of reporters, up to and including a full 'cognitive social structure' design (Krackhardt, 1987), where each respondent reports on all other ties in the network. Alongside this, the model remains computationally efficient given the use of variational inference, as opposed to a Monte Carlo approach. Our model can flexibly handle social network datasets of any realistic size, and can scale to large systems of tens of thousands of nodes by exploiting the sparsity of typical network datasets.

Results from our simulation experiments highlight that mutuality dramatically impacts inferred levels of reciprocity. Our results complement previous empirical and theoretical studies (e.g., Ready & Power, 2021; Redhead et al., Accepted), and show that the simple deterministic approach of taking the union or intersection of nominations leads to biased estimates of reciprocity. Given this, we propose a simple heuristic that is based solely on the mutuality value inferred by the model, that can be used to select the most appropriate point-estimates from an estimated posterior distribution of Y. Findings from our simulation experiments suggest that our approach results in networks that are somewhere between those produced by the union and the intersection. Generally, our approach results in lower levels of reciprocity than deterministic aggregation, because we are appropriately accounting for mutuality.

The importance of considering the core questions of (bias in) network representation—and the utility of VIMuRe—are most clearly demonstrated with our analyses of the empirical data from Karnataka (Banerjee et al., 2013) and Nicaragua (Koster, 2018). These datasets result from two very different elicitation approaches, which carry with them different potential risks for bias. The data from Karnataka provide a case where a standard name generator approach was used on a partial sample of the network, and where an upper bound of four was placed on the number of ties that could be reported. This design likely increases the chances that ties are under-reported. In contrast, the data from Nicaragua were collected using a full roster-based design on the entire sample. This approach may inflate the chances that ties are over-reported. In both empirical examples, the prospect of mutuality is salient, as reporters were asked about their roles as givers and receivers in direct succession, and there was no randomization of question order. The results of our empirical applications indicate the importance of mutuality in patterning reports within double-sampled designs. In Karnataka, mutuality values range from ~0.4 to ~0.7, and in Nicaragua they are ~0.6. These mutuality values complement the findings from our simulation experiments, which show that when mutuality is high, taking either the union or the intersection will result in inflated reciprocity values (despite treating discordant responses in very different ways).

Our findings highlight that mutuality is indeed high across a range of different relationship types, and thus the consequences of using these standard deterministic aggregation methods are obvious: a clear disparity between the resulting aggregated networks and the 'true' underlying network. The acuteness of this issue depends on the particular tie type, as we can see in the varying levels of mutuality in the Karnataka data (where mutuality is lowest for relationships that may be seen as less balanced). VIMuRe provides a promising way forward here, as it is able to measure and account for mutuality across different sampling regimes.

The empirical examples that we present further elucidate the varying 'reliabilities' of reporters—over and above the general propensity to report mutually supportive, reciprocal ties. Importantly, the contrasting results found between the two sets of empirical data reveal general issues with sampling and elicitation, about which practitioners need to be cognisant. The roster-based design used to collect the network data in Nicaragua, resulted in an average of 25 nominations for each prompt. In contrast, given the upper limit of four ties that could be reported in the Karnataka design, the average number of nominations was much lower (around two to three nominations for each prompt) for the various relationship types (see Table S2). Compounding this issue is the partial sampling procedure implemented in Karnataka. The partial sample included ~46% of the households and ~25% of residents (including children) within the sampled villages. Our findings suggest that when many nominations are of people who were not themselves reporters, there are considerable constraints on the ability to assess reliability. Moreover, our findings suggest that individuals who were named by many others are likely to be seen as 'unreliable' (see Figure S10), in part because these individuals were constrained in their ability to name more than four individuals. Generally speaking, greater coverage of the network and prompts that facilitate collection of more-complete nomination sets will permit more precise estimation of individual 'reliabilities' and, thus, more accurate network reconstruction.

Several directions are possible for future improvements to VIMuRe. Our model specifies conditional probabilities, and thus relies on pseudo-likelihood estimation for inferring the parameters. A fruitful avenue for future research is to improve this approximation by characterizing a full joint distribution of a pair of ties (Contisciani et al., 2022). Doing this may potentially solve the problem of identifying a θ_m for samples with high mutuality and, most importantly, increase the accuracy of estimating posterior distributions for Y. However, any improvement may come at the price of losing analytical tractability, or requiring less flexible approaches. We have focused here on capturing reciprocity, but this does not provide any guarantees of recovering automatically other network properties involving higher-order motifs, such as transitivity or triadic closure (see Table S2). How to adapt our model to include them is open for future research.

Alongside this, there are several other possibilities for future extensions of the VIMuRe framework that we have introduced here. First, VIMuRe takes as input a set of reported ties and we assumed that this is the only information known. However, if practitioners have access to additional information—such as covariates on nodes—this information could be incorporated into the model. Covariates could also be incorporated into models predicting reliabilities θ, and those reliabilities could vary for senders and receivers as well. For instance, one can consider a suitable prior for the reliabilities θ_m that is based upon a given covariate. It would also be straightforward to extend our model by incorporating more informative priors about the ground-truth network, Y (e.g., if the network had a known block structure). Second, many social networks are fundamentally multi-level, with nodes being nested within higher-order units (e.g., households, businesses, or schools; Lazega & Snijders, 2015). Formulating an approach to flexibly incorporate multi-level networks further remains an open and important area for extending the VIMuRe framework. Finally, our focus has been on cases where social networks are static. Investigating how to effectively adapt our model for networks evolving over time is an open avenue for future work.

In sum, there are potentially strong biases in self-reported social network data. However, the nature of multiply reported data as containing multiple sources of information about a single underlying relationship permits the application of statistical procedures that can account for such biases. VIMuRe attempts to do this by explicitly modelling mutuality—the tendency of reporters to nominate the same individuals for both directions of a tie—and estimating a reporting accuracy parameter, θ_m, for each reporter. Model estimation is performed using variational inference, leading to a fast algorithmic implementation that is scalable to large system sizes. Our study of the datasets from Karnataka and Nicaragua establishes that there is indeed important variation in reporters' 'reliability', and that people's reports seem to be driven in part by their normative expectation of relationships as balanced and reciprocal. We observe this high 'mutuality' despite very different data elicitation approaches, and see that it varies based on the type of relationship being elicited. These findings demonstrate the value of employing a tool such as VIMuRe, as it can not only give crucial insights into how social relationships are understood by individuals, but can also provide a way to account for these individual and collective biases and arrive at a more appropriate representation of the network of interest. To facilitate its usage by practitioners, we provide an open source implementation of the code online.

Acknowledgements

This paper comes out of a collaboration funded by a UKRI Economic and Social Research Council Research Methods Development Grant (ES/V006495/1). The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Martina Contisciani and Diego Baptista. Caterina De Bacco, Martina Contisciani, and Hadiseh Safdari were supported by the Cyber Valley Research Fund. Daniel Redhead, Cody T. Ross, and Richard McElreath were supported by the Department of Human Behaviour, Ecology, and Culture at the Max Planck Institute for Evolutionary Anthropology.

Data availability

All data and code used in this paper are available in the following public repository: https://github.com/latentnetworks/vimure. The data that support the findings of this study are also openly available at the following locations: The Abdul Latif Jameel Poverty Action Lab Dataverse at https://doi.org/10.7910/DVN/U3BIHX and The Royal Society Open Science's Electronic Supplementary Material at https://doi.org/10.1098/rsos.172159.

Conflict of interest: The authors declare that they have no conflict of interest.

Supplementary material

Supplementary data are available at *Journal of the Royal Statistical Society* online.

References

Adams A. M., Madhavan S., & Simon D. (2006). Measuring social networks cross-culturally. *Social Networks*, *28*(4), 363–376. https://doi.org/10.1016/j.socnet.2005.07.007

Adams J., & Moody J. (2007). To tell the truth: Measuring concordance in multiply reported network data. *Social Networks*, *29*(1), 44–58. https://doi.org/10.1016/j.socnet.2005.11.009

Amini L., Shaikh A., & Schulzrinne H. (2004). Issues with inferring internet topological attributes. *Computer Communications*, *27*(6), 557–567. https://doi.org/10.1016/j.comcom.2003.08.021

An W. (2022). You said, they said: A framework on informant accuracy with application to studying self-reports and peer-reports. *Social Networks*, *70*, 187–197. https://doi.org/10.1016/j.socnet.2021.12.006

An W., & Schramski S. (2015). Analysis of contested reports in exchange networks based on actors' credibility. *Social Networks*, *40*, 25–33. https://doi.org/10.1016/j.socnet.2014.07.002

Ball B., & Newman M. E. (2013). Friendship networks and social status. *Network Science*, *1*(1), 16–30. https://doi.org/10.1017/nws.2012.4

Banerjee A., Chandrasekhar A. G., Duflo E., & Jackson M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498–1–1236498–7. https://doi.org/10.1126/science.1236498

Bell D. C., Belli-McQueen B., & Haider A. (2007). Partner naming and forgetting: Recall of network members. *Social Networks*, *29*(2), 279–299. https://doi.org/10.1016/j.socnet.2006.12.004

Bernard H. R., Killworth P., Kronenfeld D., & Sailer L. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology*, *13*(1), 495–517. https://doi.org/10.1146/annurev.an.13.100184.002431

Besag J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 192–225. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Blei D. M., Kucukelbir A., & McAuliffe J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Borgatti S. P., Mehra A., Brass D. J., & Labianca G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

Brands R. A. (2013). Cognitive social structures in social network research: A review. *Journal of Organizational Behavior*, *34*(S1), S82–S103. https://doi.org/10.1002/job.1890

Brewer D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, *22*(1), 29–43. https://doi.org/10.1016/S0378-8733(99)00017-9

Burt R. S. (1984). Network items and the general social survey. *Social Networks*, *6*(4), 293–339. https://doi.org/10.1016/0378-8733(84)90007-8

Butts C. T. (2003). Network inference, error, and informant (in)accuracy: A Bayesian approach. *Social Networks*, *25*(2), 103–140. https://doi.org/10.1016/S0378-8733(02)00038-2

Casella G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, *39*(2), 83–87. https://doi.org/10.1080/00031305.1985.10479400

Chicco D., & Jurman G. (2020). The advantages of the Matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 1–13. https://doi.org/10.1186/s12864-019-6413-7

Contisciani M., Safdari H., & De Bacco C. (2022). Community detection and reciprocity in networks by jointly modeling pairs of edges. *Journal of Complex Networks*, *10*(4). https://doi.org/10.1093/comnet/cnac034

De Bacco C., Power E. A., Larremore D. B., & Moore C. (2017). Community detection, link prediction, and layer interdependence in multilayer networks. *Physical Review E*, *95*(4), 042317. https://doi.org/10.1103/PhysRevE.95.042317

D'haeseleer P., & Church G. M. (2004). Estimating and improving protein interaction error rates. In *Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004* (pp. 216–223). IEEE.

Eagle D. E., & Proeschold-Bell R. J. (2015). Methodological considerations in the use of name generators and interpreters. *Social Networks*, *40*, 75–83. https://doi.org/10.1016/j.socnet.2014.07.005

Eagle N., Pentland A. S., & Lazer D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, *106*(36), 15274–15278. https://doi.org/10.1073/pnas.0900282106

Fehr E., & Gächter S. (1998). Reciprocity and economics: The economic implications of Homo Reciprocans. *European Economic Review*, *42*(3–5), 845–859. https://doi.org/10.1016/S0014-2921(97)00131-1

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. https://doi.org/10.2307/2786941

Gelman A., Meng X.-L., & Stern H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–760.

Gelman A., Stern H. S., Carlin J. B., Dunson D. B., Vehtari A., & Rubin D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Givens C. R., & Shortt R. M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, *31*(2), 231–240. https://doi.org/10.1307/mmj/1029003026

Heider F. (1958). *The psychology of interpersonal relations*. Wiley.

Hobson E. A., Silk M. J., Fefferman N. H., Larremore D. B., Rombach P., Shai S., & Pinter-Wollman N. (2021). A guide to choosing and implementing reference models for social network analysis. *Biological Reviews*, *96*(6), 2716–2734. https://doi.org/10.1111/brv.v96.6

Hoff P., Fosdick B., Volfovsky A., & Stovel K. (2013). Likelihoods for fixed rank nomination networks. *Network Science*, *1*(1), 253–277. https://doi.org/10.1017/nws.2013.17

Hoff P. D., Raftery A. E., & Handcock M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098. https://doi.org/10.1198/016214502388618906

Holland P. W., Laskey K. B., & Leinhardt S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137. https://doi.org/10.1016/0378-8733(83)90021-7

Holt-Lunstad J., Smith T. B., Baker M., Harris T., & Stephenson D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, *10*(2), 227–237. https://doi.org/10.1177/1745691614568352

Jackson M. O. (2021). Inequality's economic and social roots: The role of social networks and homophily. *SSRN Scholarly Paper ID 3795626*, Social Science Research Network, Rochester, NY.

Jackson M. O., Rodriguez-Barraquer T., & Tan X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, *102*(5), 1857–1897. https://doi.org/10.1257/aer.102.5.1857

Karrer B., & Newman M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, *83*(1), 016107. https://doi.org/10.1103/PhysRevE.83.016107

Kenny D. A., & La Voie L. (1984). The social relations model. In *Advances in experimental social psychology* (Vol. *18*, pp. 141–182). Academic Press.

Killworth P., & Bernard H. (1976). Informant accuracy in social network data. *Human Organization*, *35*(3), 269–286. https://doi.org/10.17730/humo.35.3.10215j2m359266n2
Koster J. (2018). Family ties: The multilevel effects of households and kinship on the networks of individuals. *Royal Society Open Science*, *5*(4), 172159. https://doi.org/10.1098/rsos.172159
Krackhardt D. (1987). Cognitive social structures. *Social Networks*, *9*(2), 109–134. https://doi.org/10.1016/0378-8733(87)90009-8
Krackhardt D., & Kilduff M. (1990). Friendship patterns and culture: The control of organizational diversity. *American Anthropologist*, *92*(1), 142–154. https://doi.org/10.1525/aa.1990.92.issue-1
Kristiansen S. (2004). Social networks and business success: The role of subcultures in an African context. *American Journal of Economics and Sociology*, *63*(5), 1149–1171. https://doi.org/10.1111/ajes.2004.63.issue-5
Lazega E., & Snijders T. A. B. (Eds.) (2015). *Multilevel network analysis for the social sciences: Theory, methods and applications*. Springer.
Lazer D., Hargittai E., Freelon D., Gonzalez-Bailon S., Munger K., Ognyanova K., & Radford J. (2021). Meaningful measures of human society in the twenty-first century. *Nature*, *595*(7866), 189–196. https://doi.org/10.1038/s41586-021-03660-7
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. https://doi.org/10.1016/j.socnet.2018.05.005
Lin N. (2002). *Social capital: A theory of social structure and action*. Cambridge University Press.
Lungeanu A., McKnight M., Negron R., Munar W., Christakis N. A., & Contractor N. S. (2021). Using Trellis software to enhance high-quality large-scale network data collection in the field. *Social Networks*, *66*, 171–184. https://doi.org/10.1016/j.socnet.2021.02.007
Marin A. (2004). Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Social Networks*, *26*(4), 289–307. https://doi.org/10.1016/j.socnet.2004.06.001
Marineau J. E., Labianca G. J., 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. https://doi.org/10.1016/j.socnet.2018.01.006
Marsden P. V. (1990). Network data and measurement. *Annual Review of Sociology*, *16*(1), 435–463. https://doi.org/10.1146/soc.1990.16.issue-1
Marsden P. V. (2003). Interviewer effects in measuring network size using a single name generator. *Social Networks*, *25*(1), 1–16. https://doi.org/10.1016/S0378-8733(02)00009-6
Marsden P. V. (2005). Recent developments in network measurement. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 8–30), No. 27 in Structural Analysis in the Social Sciences. Cambridge University Press.
Molm L. D. (2010). The structure of reciprocity. *Social Psychology Quarterly*, *73*(2), 119–131. https://doi.org/10.1177/0190272510369079
Morris C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American statistical Association*, *78*(381), 47–55. https://doi.org/10.1080/01621459.1983.10477920
Newcomb T. M. (1961). The acquaintance process as a prototype of human interaction. In T. M. Newcomb (Ed.), *The acquaintance process* (pp. 259–261). Holt, Rinehart & Winston.
Newman M. E. J. (2018a). Estimating network structure from unreliable measurements. *Physical Review E*, *98*(6), 062321. https://doi.org/10.1103/PhysRevE.98.062321
Newman M. E. J. (2018b). Network structure from rich but noisy data. *Nature Physics*, *14*(6), 542–545. https://doi.org/10.1038/s41567-018-0076-1
Nolin D. A. (2008). *Food-sharing networks in Lamalera, Indonesia: Tests of adaptive hypotheses* [Ph.D. thesis]. University of Washington.
Nolin D. A. (2010). Food-sharing networks in Lamalera, Indonesia: Reciprocity, kinship, and distance. *Human Nature*, *21*(3), 243–268. https://doi.org/10.1007/s12110-010-9091-3
Park P. S., Blumenstock J. E., & Macy M. W. (2018). The strength of long-range ties in population-scale social networks. *Science*, *362*(6421), 1410–1413. https://doi.org/10.1126/science.aau9735
Peixoto T. P. (2018). Reconstructing networks with unknown and heterogeneous errors. *Physical Review X*, *8*(4), 041011. https://doi.org/10.1103/PhysRevX.8.041011
Perkins J. M., Subramanian S. V., & Christakis N. A. (2015). Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. *Social Science & Medicine*, *125*, 60–78. https://doi.org/10.1016/j.socscimed.2014.08.019
Power E. A. (2017). Social support networks and religiosity in rural South India. *Nature Human Behaviour*, *1*(3), 0057. https://doi.org/10.1038/s41562-017-0057
Pustejovsky J. E., & Spillane J. P. (2009). Question-order effects in social network name generators. *Social Networks*, *31*(4), 221–229. https://doi.org/10.1016/j.socnet.2009.06.001

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. https://doi.org/10.1086/696018

Ready E., & Power E. A. (2021). Measuring reciprocity: Double sampling, concordance, and network construction. *Network Science*, *9*(4), 387–402. https://doi.org/10.1017/nws.2021.18

Redhead D., McElreath R., & Ross C. T. (Accepted). Reliable network inference from unreliable data: A tutorial on latent network modeling using STRAND. *Psychological Methods*. Advance online publication. https://psyarxiv.com/mkp2y/

Redhead D., & Power E. A. (2022). Social hierarchies and social networks in humans. *Philosophical Transaction of the Royal Society B*, *377*(1845), 20200440. https://doi.org/10.1098/rstb.2020.0440

Redhead D., & von Rueden C. R. (2021). Coalitions and conflict: A longitudinal analysis of men's politics. *Evolutionary Human Sciences*, *3*, E31. https://doi.org/10.1017/ehs.2021.26

Robbins H. E. (1955). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Probability and Statistics* (pp. 157–164).

Robins G., Pattison P., Kalish Y., & Lusher D. (2007). An introduction to exponential random graph ($p*$) models for social networks. *Social Networks*, *29*(2), 173–191. https://doi.org/10.1016/j.socnet.2006.08.002

Ross C. T., & Redhead D. (2022). DieTryin: An R package for data collection, automated data entry, and post-processing of network-structured economic games, social networks, and other roster-based dyadic data. *Behavior Research Methods*, *54*(2), 611–631. https://doi.org/10.3758/s13428-021-01606-5

Safdari H., Contisciani M., & De Bacco C. (2021). Generative model for reciprocity and community detection in networks. *Physical Review Research*, *3*(2), 023209. https://doi.org/10.1103/PhysRevResearch.3.023209

Sewell D. K. (2019). Latent space models for network perception data. *Network Science*, *7*(2), 160–179. https://doi.org/10.1017/nws.2019.1

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. https://doi.org/10.1016/j.socnet.2016.08.008

Simpson B., Markovsky B., & Steketee M. (2011). Power and the perception of social networks. *Social Networks*, *33*(2), 166–171. https://doi.org/10.1016/j.socnet.2010.10.007

Simpson C. R. (2022). On the structural equivalence of coresidents and the measurement of village social structure. *Social Networks*, *69*, 55–73. https://doi.org/10.1016/j.socnet.2020.02.010

Smith K. P., & Christakis N. A. (2008). Social networks and health. *Annual Review of Sociology*, *34*(1), 405–429. https://doi.org/10.1146/soc.2008.34.issue-1

Sosa J., & Rodríguez A. (2021). A latent space model for cognitive social structures data. *Social Networks*, *65*, 85–97. https://doi.org/10.1016/j.socnet.2020.12.002

Sprinzak E., Sattath S., & Margalit H. (2003). How reliable are experimental protein–protein interaction data? *Journal of Molecular Biology*, *327*(5), 919–923. https://doi.org/10.1016/S0022-2836(03)00239-0

Titterington D., Smith A., & Makov U. (1985). *Statistical analysis of finite mixture distributions*. Wiley.

von Rueden C., Redhead D., O'Gorman R., Kaplan H., & Gurven M. (2019). The dynamics of men's cooperation and social status in a small-scale society. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1908), 20191367. https://doi.org/10.1098/rspb.2019.1367

Warner R. M., Kenny D. A., & Stoto M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, *37*(10), 1742–1757. https://doi.org/10.1037/0022-3514.37.10.1742

Young J.-G., Cantwell G. T., & Newman M. E. J. (2021). Bayesian inference of network structure from unreliable data. *Journal of Complex Networks*, *8*(6), cnaa046. https://doi.org/10.1093/comnet/cnaa046

Yousefi-Nooraie R., Marin A., Hanneman R., Pullenayegum E., Lohfeld L., & Dobbins M. (2019). The relationship between the position of name generator questions and responsiveness in multiple name generator surveys. *Sociological Methods & Research*, *48*(2), 243–262. https://doi.org/10.1177/0049124117701484