LATENT NETWORK MODELS TO ACCOUNT FOR NOISY, MULTIPLY-REPORTED SOCIAL NETWORK DATA

Caterina De Bacco^{1, *}, Martina Contisciani¹, Jonathan Cardoso-Silva², Hadiseh Safdari¹, Diego Baptista¹, Tracy Sweet³, Jean-Gabriel Young⁴, Jeremy Koster^{5,6}, Cody T. Ross⁷, Richard McElreath⁷, Daniel Redhead^{7, *}, and Eleanor A. Power^{2,*}

¹Max Planck Institute for Intelligent Systems, Cyber Valley, Tuebingen 72076, Germany

²Department of Methodology, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK

³Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20742, USA

⁴Department of Mathematics and Statistics and Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405

⁵Department of Anthropology, University of Cincinnati, Cincinnati, OH 45221, USA

⁶National Science Foundation, Alexandria, VA 22314, USA

⁶Department of Human Behaviour, Ecology & Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany

^{*}Corresponding authors: caterina.debacco@tuebingen.mpg.de, daniel_redhead@eva.mpg.de & 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 75 Indian villages collected with a name-generator design, and a Nicaraguan community collected with a roster-based 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 latent variable · information accuracy · variational inference · social network analysis · network model

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 individuals' perceptions of their relationships, have been found to be important for their material wealth (Jackson, 2021), social position and welfare (Lin, 2002; Redhead and Power, Accepted), and health and well-being (Perkins et al., 2015; Holt-Lunstad 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; Park et al., 2018; Lazer et al., 2021), 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 (Warner et al., 1979; Marsden, 2005; Ross and Redhead, 2021).

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 (Krackhardt, 1987; Freeman, 1992). Questions may be framed around more qualitative sentiments towards others—such as 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 and 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 and 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 (Killworth and Bernard, 1976; Bernard et al., 1984). 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% - 59% of their interactions across a 24 hour 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, and the duration of a given relationship (Bell et al., 2007). Both theoretical studies and empirically observed patterns of nomination also 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 and Newman, 2013; Marin, 2004; Redhead et al., 2021; Shakya et al., 2017; Marineau et al., 2018). The order in which questions appear within a survey, and the mode of elicitation, may further influence responses (Pustejovsky and Spillane, 2009; Eagle and Proeschold-Bell, 2015). That is, respondents have been shown to become fatigued, and report fewer relationships, when asked several name generator questions (Yousefi-Nooraie et al., 2019). And, responses can also vary between interviewers, based in part on their attributes and their dynamic with the interviewee (Marsden, 2003; Lungeanu et al., 2021).

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 - 90% (Adams and 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 sampling, double sampling provides two perspectives on all relationships within the network, as both the giver and receiver have an opportunity to name their partner for one of the two prompts. A recent survey of double-sampled network data has suggested that concordance between reporters is low, with an overall average of 10% agreement (Ready and 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, and has highlighted the individual differences that may guide respondents' perceptions of their relationships and the relationships of others (see Brands, 2013, for a review).

These low levels of 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 multiple prompts entailed in double sampling may lead to an inflation of observed reciprocity, driven primarily by people's propensity of naming of the same individuals across both prompts (Ready and Power, 2021), which we here term "mutuality." Overall, the low levels of concordance found in multiply-reported data raises the question of how to integrate such imperfect reports elicited from respondents with appropriate statistical models.

LATENT NETWORK MODELS TO ACCOUNT FOR NOISY, MULTIPLY-REPORTED SOCIAL NETWORK DATA

To examine the individual biases that shape self-reports of ties, and estimate the effect of mutuality on core properties of a network, we advance a statistical estimation technique that is able to combine multiply reported network data while accounting for the variable "reliability" of respondents. We validate our model using simulated data, to determine whether we are able to recover the true underlying network, mutuality, and individual reporters' reliabilities. Finally, we evaluate our model with two empirical datasets, both with double-sampled questions, one based on a "name generator" approach and the other based on the roster method. We conclude by discussing our findings and outlining possible extensions to the model.

1.1 Related Work

In the social sciences, simple deterministic rules are often used to aggregate potentially multiple reports on what should nominally be the same relationship (Krackhardt, 1987; Lee and Butts, 2018). With 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 and Power, 2018). Alternatively, it could be assumed that only relationships that are mutually recognised are salient; with such an expectation the intersection of the two name generators would be the implied choice (e.g., Krackhardt and Kilduff, 1990). These aggregation rules rest on simple but strong expectations, and presume consistency in how reporters respond to these questions. This, paired with the fact that standard statistical tools used in the social sciences (e.g., exponential random graph models (Robins et al., 2007), or social relations models (Snijders and Kenny, 1999)) assume that reported ties are a "true" representation of the network, can potentially lead to serious misrepresentations of the social relations of interest.

In recent years, several statistical methods have been proposed to resolve discordant reports, in social network analysis (Holland et al., 1983; Killworth and Bernard, 1976; Kenny and La Voie, 1984) but also in other fields like systems engineering (Amini et al., 2004), the biological sciences (Sprinzak et al., 2003; D'haeseleer and Church, 2004; Hobson et al., 2020), 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 base on each reporter's assigned credibility (An and Schramski, 2015).

Among this vast literature, most relevant to the work presented here are Bayesian approaches to network estimation (Butts, 2003). They rely on an explicit generative models for the reports, in which the reports are viewed as giving indirect and imperfect information about an unobserved network,. Virtually all of these generative models are formulated as finite mixtures (Titterington et al., 1985), where responses about a particular tie have different distributions depending on the presence or absence of that tie. Formalized thus, these models can often be estimated with efficient algorithms such as expectation-maximization algorithms (Newman, 2018a,b), exact Gibbs samplers (Butts, 2003; Young et al., 2021b), or Hamiltonian Monte Carlo samplers applied to closed marginal distributions (Young et al., 2021a,c; Redhead et al., 2021).

Generative approaches are fairly flexible and can readily be adapted to diverse experimental and observational situations. For example, by allowing the distributions of the data to vary from tie to tie, mixture models of networks have been shown to accommodate surveys where respondents have various degrees of trustworthiness (Butts, 2003), some of the data is missing (Peixoto, 2018, 2019) and more (see Redhead et al., 2021). Further, important social phenomena driving tie formation, such as triadic closure, reciprocity and group structure, can be integrated through specification of priors on the network (such as with stochastic block models; Peixoto, 2018, 2021) or latent space models (Butts, 2003)), or by suitably incorporating them into a likelihood distribution (Safdari et al., 2021a,b; Contisciani et al., 2021).

Given the modelling and computing advantages provided by generative mixture models, we introduce a model complementing the previous work outlined above. We go beyond existing approaches by introducing a model that simultaneously handles multiply-reported ties and weighted reports, while allowing individuals to vary in reliability. For technical reasons, the resulting model is no longer a *finite* mixture (unlike previous work), which in turns means most of the standard statistical methods do not apply. As we will show, this issue can be resolved by introducing an efficient variational inference procedures.

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 existing 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 $Y_{ij} \in \{0, 1\}$ indicate whether a tie $i \rightarrow j$ exists. 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 $i \rightarrow j$.

We assume that each reporter 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 reports 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), one for each direction of the interaction.

For instance, m can nominate who she gives advice to (giving) and who she goes to for advice (receiving). In this case $m \in \{i, j\}$ for a tie between i and j, and we distinguish the reported data of each tie by X_{ijm} and X_{jim} , depending on it being given or received. Generally, we assume the reported data X to be asymmetric in (i, j). While we gave an example for ties of type *advice*, the model applies for any type of tie.

One of the main objectives of our model is to estimate the structure of the unobserved network, Y, from the reported data, X. 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 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., 2021a), and makes the 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 nominations reduces 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, 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) certain ties, and this may depend on tie type. To account for these biased reports we assign a "reliability" parameter, θ_m , to each reporter, m, specific to the tie type given in input. 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 relationships, X_{ijm} and X_{jim} . We term this "mutuality," to keep the notion distinct from the standard dyadic reciprocity (henceforth termed reciprocity) in the true unobserved network Y.

Bringing all of these modeling consideration together, we posit that the expected value of the data is:

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

where $\eta \ge 0$ is the mutuality parameter. Mutuality intervenes as an additive and positive contribution to the mean number of reported ties X_{ijm} , capturing the possible increasing probability of observing a tie, given that we observed 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 and regulates the contribution of Y in determining X.

From this we notice how, for a given set of $\lambda_k > 0$, reporters with high θ_m tend to exaggerate their reports, while reporters with smaller values tend to nominate fewer individuals. Regardless of the reporter's 'reliability', the existence of a tie X_{jim} in one direction increases the probability that the tie X_{ijm} also exists in the opposite direction when $\eta > 0$. This also implies that it may not be possible to identify the reliability of reporters with high mutuality 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 on the opposite direction, no further information about the reporter (e.g., reliability) needs to be estimated. To form a likelihood for the observed data that can accommodate various network and report structures, in particular directed and weighted reports and networks, we write

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

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

In addition to specifying the likelihood as in Eq. (2), we adopt a Bayesian approach and assume priors for the parameters and the unobserved Y. To maximise the flexibility of our model, we allow for different types of values for X and Y, leading to the formulation:

$$P(Y_{ij}; p_{ij}) = \prod_{k} (p_{ij,k})^{Y_{ij,k}} ,$$
(3)

with $\sum_k p_{ij,k} = 1$. The resulting model can thus accommodate, for example, a binary network Y and weighted reports X. We then consider Gamma priors for the remaining parameters, as they are defined for positive real numbers, and they 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 use a pseudo-likelihood approximation as in Safdari et al. (2021a):

$$P\left(\left\{X_{ijm}\right\}_{m} | \left\{X_{jim}\right\}_{m}, Y_{ij}, \lambda, \left\{\theta_{m}\right\}_{m}, \eta\right) = \prod_{m} P(X_{ijm} | X_{jim}, Y_{ij}, \lambda, \theta_{m}, \eta)$$
$$= \prod_{k} \left[P(Y_{ij} = k) \prod_{m} P(X_{ijm} | X_{jim}, Y_{ij} = k, \lambda_{k}, \theta_{m}, \eta)\right]^{Y_{ij,k}}.$$
(4)

The full posterior is then:

$$P(Y,\lambda,\theta,\eta|X) \propto P(X|Y,\lambda,\theta,\eta)P(Y)P(\lambda)P(\theta)P(\eta)$$

$$= \prod_{i,j} P\left(\left\{X_{ijm}\right\}_{m} | \left\{X_{jim}\right\}_{m}, Y_{ij}, \lambda, \left\{\theta_{m}\right\}_{m}, \eta\right)P(Y_{ij}; p_{ij}) \qquad (5)$$

$$\prod_{i,k} P(\lambda_{k}; a_{k}, b_{k}) \prod_{i,m} P(\theta_{m}; \alpha_{m}, \beta_{m})P(\eta; c, d)$$

$$=: \mathcal{L}(\lambda, \theta, \eta, Y) \quad , \qquad (6)$$

where the proportionality is due to 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 finds 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 the marginal posterior distribution of each parameter while holding the others fixed. We call the resulting algorithm VIMuRe, for Variational Inference for Multiply-Reported data. Pseudo-code for the resulting 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.

Algorithm 1: VIMuRe

Input: Data X, Model \mathcal{L} , Variational family q. Initialize the variational parameters γ , ϕ , ρ , ν to the priors with a small random offset. while change in ELBO is above a threshold do For each pair of nodes such that $X_{iim} > 0$, update the multinomials: $\hat{z}_{mk}^{1} \propto \exp\left\{\Psi(\gamma_{m}^{shape}) - \log\gamma_{m}^{rate} + \Psi(\phi_{k}^{shape}) - \log\phi_{k}^{rate}\right\}$ $\hat{z}_{ijm}^2 \propto \exp\left\{\Psi(\nu^{shape}) - \log\nu^{rate} + \log X_{jim}\right\} = X_{jim} \exp\left\{\Psi(\nu^{shape}) - \log\nu^{rate}\right\}$ where the proportionality is such that $\hat{z}_{mk}^1 + \hat{z}_{ijm}^2 = 1$. For each reporter, update the reliability parameters: $\gamma_m^{shape} = \alpha_m + \sum_{i \ i \ k} R_{ijm} \, \rho_{ij,k} \, X_{ijm} \, \hat{z}_{mk}^1$ $\gamma_m^{rate} = \beta_m + \sum_{i \ i \ k} R_{ijm} \rho_{ij,k} \frac{\phi_k^{shape}}{\phi_k^{rate}} \,.$ For each possible value k of Y_{ij} , update the parameters: $\phi_k^{shape} = a_k + \sum_{i,j,m} R_{ijm} \rho_{ij,k} X_{ijm} \hat{z}_{mk}^1$ $\phi_k^{rate} = b_k + \sum_{i \ i \ m} R_{ijm} \ \rho_{ij,k} \frac{\gamma_m^{shape}}{\gamma_m^{rate}}$ and: $\rho_{ij,k} \propto \exp\left\{\log p_{ij,k} + \sum R_{ijm} \left(X_{ijm} \hat{z}^{1}_{mk} \mathbb{E}_{q(\lambda_{k})} \left[\log \lambda_{k}\right]\right) - \frac{\phi^{shape}_{k}}{\phi^{rate}_{k}} \sum R_{ijm} \frac{\gamma^{shape}_{m}}{\gamma^{rate}_{m}}\right\} .$ Update the mutuality parameters: $\nu^{shape} = c + \sum_{i \ i \ k} \rho_{ij,k} \sum_{m} R_{ijm} X_{ijm} \hat{z}_{ijm}^2$ $\nu^{rate} = d + \sum_{i \neq m} R_{ijm} X_{jim} .$

end

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

In detail, we first generate the network Y either with a stochastic block model (SBM, Holland et al., 1983), a degreecorrected stochastic block model (DC-SBM, Karrer and Newman, 2011) or a probabilistic model with reciprocity (CRep, Safdari et al., 2021a). We then generate the observed X given a fixed reliability, mutuality and reciprocity parameters and the generated network, collectively denoted by $\Theta = (Y, \theta, \lambda, \eta)$. We follow the approach described in Safdari et al. (2021a), 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 Section S1.3 for details.

In these simulations we examine our ability to recover (i) the underlying network Y and (ii) the individual reliabilities θ_m . We generated 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 (a) over-reporters or (b) 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 show model performance in difficult cases where the proportion of unreliable reporters are increasing. The third scenario, (c), is more realistic. In this setting, we have both over- and under-reports, and we drawn θ_m from a Gamma distribution, providing a broader range of values. We vary the difference between λ_1 and λ_0 , such that the smaller this difference, the noisier the problem gets, and thus the harder the inference tasks.

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; (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.



Figure 1: Synthetic networks with over- or under-reporters: estimating underlying network Y. Synthetic networks with N = 100 nodes and M = 100 reporters generated with the benchmark generative model proposed above 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 are calculated over ten independent synthetic networks. The accuracy of the estimate of the underlying network Y is measured with the F1-score. This measure goes from 0 to 1, where 1 indicates perfect matching.

4.1 Results

We use the F1-score to assess the ability of our model to recover Y (which is binary in our experiments). It is defined as the harmonic mean of precision (fractions of inferred ties that actually exists) and recall (fractions of existing ties found by the method). This choice is chiefly motivated by the fact that we have unbalanced data, since much fewer ties than possible tend to exists in empirical network as well as in our experiments.

In the two extreme scenarios, where there are only over- or under-reporters, our model recovers the unobserved network Y better than the union or intersection of X. The performance of our model is also more robust as the number of unreliable reporters and 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 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' reliability, and this results in higher and more robust performance. However, when θ_{ratio} become 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.



Figure 2: Synthetic networks with over- and under-reporters. Synthetic networks with N = 100 nodes and M = 100 reporters generated with the benchmark generative model proposed above by varying the difference between λ_0 and λ_1 . The three columns represent networks generated with no (left), medium (center) and high (right) mutuality effect η . The results are averages and standard deviations over twenty independent synthetic networks, and the accuracy over the underlying network Y is measured with F1-score.

Performance is more nuanced when we consider the more realistic experiment, with a broad range of reporters' reliabilities. F1 scores are lower than in the previous experiments in general, and recovering the ground truth is particularly harder 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 non-zero inputs of Y tend to make the same contribution in determining X; thus, it becomes more difficult to distinguish between reports about ties and absent ties. These experiments also further confirm what we observed in the previous experiments, that the harder regime is with 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 having a less structured X. In these experiments, both versions of our model and the union baseline perform similarly while the intersection baseline performs much more poorly. However, the performance gap decreases as mutuality increases and more ties are reported.

Once we have an estimate \hat{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 \hat{Y} , a foundational feature of many social relations (Fehr and 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 $\hat{Y}_{ij} = 1$ 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 $\hat{Y}_{ij} = 1$ when $\rho_{ij,k=1} \ge t_p$, and keep track of the best t_{ρ}^* , for which reciprocity in the inferred network most closely matched 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:

$$t_o^* = 0.54 \,\eta_{est} - 0.01. \tag{7}$$

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 the ground truth when compared to baseline methods in most scenarios, as it can be seen in Figure S4. These results mean that despite the small gap on the value of F1-score, VIMuRe_T may be able to provide a better estimate of structural properties on Y.

As a final test, we also show that VIMuRe can recover the community structure of a network even when it is measured noisily. To this end, we use a latent network that has planted communities, and generate reports as before. We then estimate the network, and finally recover communities using a probabilistic generative model with latent variables



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



Figure 4: Synthetic networks with N = 100 nodes and M = 100 reporters generated with the benchmark generative model proposed above with $\lambda_1 - \lambda_0 = 1.0$ and planted reciprocity values around ≈ 0.2 on the ground truth network. The four sub-plots represent networks generated with low (top left), medium (top right) to increasingly high (bottom left and right) mutuality effect η . The box plots are distributions of the reciprocity on Y, over twenty samples of synthetic networks.

(De Bacco et al., 2017), and run it on the estimated \hat{Y} using each method. Figure 5 shows the result of this experiment and further confirms this in relation to recovering the ground truth communities used to generate Y, which is also relevant to many potential applications. VIMuRe is more robust than the other approaches that have been considered in relation to F1-score across the 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 others, where the two true communities are more mixed.



Figure 5: Synthetic networks with both over- and under-reporters: recovering community structure. Synthetic networks with N = 100 nodes and M = 100 reporters generated with the benchmark generative model proposed above with $\lambda_1 - \lambda_0 = 1.0$. (left) The results are averages and standard deviations over twenty samples of synthetic networks, generated by varying the mutuality parameter η . The accuracy in recovering the community structure is measured with the F1-score using the inferred membership vectors. (right) Example of a resulting partition from a synthetic network generated with $\eta = 0.4$. The "ground truth" is the partition used to generate Y, and Y_{GT} stands for the partition found by using the true Y.

To summarize, our simulation experiments suggest that working with a generative model with latent variables results in more robust estimates of the underlying and true network, Y, in comparison to deterministic baselines (such as union and intersection). Furthermore, our model provides an estimate of reporter reliabilities, which can provide additional insights about the networks in and of themselves. In addition, we notice better performance of our model when we include the mutuality parameter η . In particular, VIMuRe_T shows better results in estimating reciprocity, specifically as people's propensity to report mutuality in their relationships increases. In the application on real-data we use VIMuRe with the mutuality parameter.

5 Analysis of social support networks in Karnataka

We apply our modelling approach to a dataset of social support networks 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 they would give or receive advice from/to (*Advice*), who they would borrow or lend a small amount of money from/to (*Money*), who they would go to or receive as visitors (*Visit*), and who they would borrow or lend kerosene and rice from/to (*Household Items – "HH Items" in the plots*). In the past, these data have been studied by aggregating multiple household respondents answers and taking the union of the double-sampled questions (Jackson et al., 2012; Banerjee et al., 2013).

Our results suggest that the reciprocity values in these networks are in fact lower than what would be obtained by these simpler approaches, as shown in Figure 6 and illustrated for one village and tie type in Figure 7. While we do not have ground truth values in this 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



Figure 6: Reciprocity on Karnataka networks. The box plots show the distribution over the 75 networks of the reciprocity measured on the inferred \hat{Y} . Each column is a different tie type, as written on the figure title.

union and intersection significantly overestimate the reciprocity on Y, whereas VIMuRe identifies the correct range of values. These results suggest that reciprocity would likely be overestimated in double-sampling network data when the relationships elicited 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 (0.296 ± 0.047), followed by "Money" (0.346 ± 0.050), and then the "Visit" (0.367 ± 0.047) and "Household items" (0.378 ± 0.042) layers exhibiting the highest 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 broadly align with expectations: 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 (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 reporters' "reliabilities" are distributed in these networks. Since the mutuality in these graphs is high ($n_{est} \ge 0.4$), it is very common that reporters repeat the same names across the different name generators and therefore it is expected that individual "reliability" will play a smaller role in determining the latent social network. Recalling Eq. (1) and the following discussion, 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%. These should not be interpreted as under-reporting, as a small θ_m in this case is just a signal of a predictable reporter's behavior due to 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, even after having discounted those individuals who have low θ_m for the reasons discussed above (see Figure S5). This is consistent with prior literature (Butts, 2003) that suggests that reporters are reasonably accurate when reporting ties, but quite inaccurate when reporting non-ties (as when under-reporting) and 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 notice a negative correlation between θ_m and the in-degree of ties reported by others involving m (see Figure S7). In particular, reporters nominated by many others (some by more than 20 people) could only nominate up to four among these; such reporters 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 S5). To explore further, we assess whether reporters are consistent in their reliabilities across the tie types by examining the pairwise distance using the Wassertein distance (a metric for measuring distances between two distributions; Givens and



(a) Union (recip. = 0.93)

(b) Intersection (recip. = 0.88) (c) VIMuRe (recip. = 0.49)

Figure 7: Example of networks estimated by baseline methods and VIMuRe for one Karnataka village (tie type "Visit").

Shortt, 1984) between each set of tie types (see Figure S6). We see some telling patterns by looking at the consistency of "Advice" with the other tie types: while reporters are most consistent between that and "Money," they are least consistent between that and "Household Items" and "Visit."

6 Analysis of Nicaragua data

To further highlight the broad applicability and benefits of our modelling approach, we apply VIMuRe to data from a horticulturalist community in Nicaragua (see Koster, 2018, for more detail on the population and measurement instruments). These data were collected through 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 that captured relationships of 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, 2020). 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 while others much fewer. It is, therefore, reasonable to have the priors on θ_m reflecting 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 allows us to obtain a wider range of reliabilities. This is particularly important as we can distinguish possible exaggerators more distinctively than when using the same prior for all reporters.



Figure 8: 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 $\hat{\theta}$ is printed on top. Each slice of the pie is one tie reported in one the six possible ways, represented by the colors. In this example we consider reporters from the Nicaragua dataset.

As with our previous examples, applying VIMuRe produces estimates of a network 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 $\eta_{est} = 0.610$ and reciprocity was 0.278. These results align closely with the values found in the Karnataka data. Indeed, it is particularly close to the values for the "Money" tie type in the Karnataka data. The "tangible support" considered here would potentially be closest to that, or the "Household Items" tie type, which had higher values for both reciprocity and mutuality.

In contrast to the Karnataka dataset, where survey design may contribute to under-reporting, the roster-based design may make 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 Karnataka study (see Table S2 for full network statistics). As can be seen in Figure 8, we 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 but these ties are not confirmed by the 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 and these ties are confirmed, but this person does not confirm several ties that others report. The distribution of reliability for reporters in this dataset can be seen in Figure S8.

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. 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 advance statistical tools for more nuanced understanding of individuals' unique potential to misrepresent or misreport their relationships. 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., 2021a), there has been limited formal analysis of the impact that "mutuality" has on network inference (but see, Redhead et al., 2021).

We have focused our attention on cases where multiple reporters are able to provide information on any given relationship, providing crucial empirical data that could be used to assess the biases under investigation here. 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. Within our framework, VIMuRe takes as input the potentially biased, imperfect survey responses and uses these to estimate a "true" latent network, as well as any 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 (while accounting for mutuality).

The model that we have introduced strongly departs from standard approaches for dealing with double-sampled network data in the social sciences, which involve researchers simply taking 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., 2021a) that are applicable to social networks focus on the single-sampled case–with the exception of Redhead et al. (2021), which is applicable only to double-sampled networks. While we have highlighted double-sampled network data, 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 is computationally efficient given the use of variational inference, as opposed to a Monte Carlo approach. Given this, 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 the network datasets.

Results from our simulation experiments highlight that mutuality dramatically impacts inferred levels of reciprocity. Our results complement previous theoretical and empirical studies (Ready and Power, 2021; Redhead et al., 2021), and show that the simple deterministic rules of the union and intersection of nominations lead to over-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 much lower levels of reciprocity than such 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 may likely increase the chances of under-reporting of ties. In contrast, the data from Nicaragua were collected using a full roster-based design on the entire sample. This approach may inflate the chances of over-reporting of ties. 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. This complements 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 network and the "true" underlying network. How acute this issue is depends on the particular tie type, as we can see in the varying levels 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 a broad range of 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 about sampling and elicitation that practitioners need to be cognisant of. Given the roster-based design used to collect the network data from Nicaragua, the average number of nominations that were given for each prompt was roughly 25. 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 (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 $\sim 46\%$ of the households and $\sim 25\%$ of residents (including children). 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 S6), 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 result in more nominations will facilitate 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., 2021). 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 possibilities for future extensions of the VIMuRe framework that we have introduced here. First, VIMuRe takes as input a set of reported ties generated flexibly from a variety of data collection routines and we assumed that this is the main information known. However, if practitioners have access to additional information—such as covariates on nodes—this could be incorporated into the model. 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 and Snijders, 2015). Formulating an approach to flexibly incorporate multi-level networks further remains an open and important area for extending our 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, multiply-reported social network data clearly highlight the potential for biases in self-report data, but also provide that material which, if properly mobilized, can help to account for those biases. VIMuRe attempts to do so by explicitly modelling mutuality, the tendency of reporters of nominating the same individuals for both directions of a tie, and estimating a parameter θ_m for each reporter, which can be related to their tendency to report various amounts of ties. Inference is based on 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 and Code

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.

References

- Adams, A. M., Madhavan, S. and Simon, D. (2006) Measuring social networks cross-culturally. *Social Networks*, **28**, 363–376.
- Adams, J. and Moody, J. (2007) To tell the truth: Measuring concordance in multiply reported network data. *Social Networks*, **29**, 44–58.
- Amini, L., Shaikh, A. and Schulzrinne, H. (2004) Issues with inferring internet topological attributes. *Computer Communications*, **27**, 557–567.
- An, W. and Schramski, S. (2015) Analysis of contested reports in exchange networks based on actors' credibility. *Social Networks*, **40**, 25–33.
- Ball, B. and Newman, M. E. (2013) Friendship networks and social status. *Network Science*, 1, 16–30.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2013) The diffusion of microfinance. *Science*, **341**, 1236498–1–7.
- Bell, D. C., Belli-McQueen, B. and Haider, A. (2007) Partner naming and forgetting: Recall of network members. *Social Networks*, **29**, 279–299.

- Bernard, H. R., Killworth, P., Kronenfeld, D. and Sailer, L. (1984) The problem of informant accuracy: the validity of retrospective data. *Annual Review of Anthropology*, 13, 495–517.
- Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2017) Variational inference: A review for statisticians. *Journal of the American Statistical Association*, **112**, 859–877.
- Borgatti, S. P., Mehra, A., Brass, D. J. and Labianca, G. (2009) Network analysis in the social sciences. *Science*, **323**, 892–895.
- Brands, R. A. (2013) Cognitive social structures in social network research: A review. *Journal of Organizational Behavior*, **34**, S82–S103.
- Brewer, D. D. (2000) Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, **22**, 29–43.
- Burt, R. S. (1984) Network items and the General Social Survey. Social Networks, 6, 293–339.
- Butts, C. T. (2003) Network inference, error, and informant (in)accuracy: a Bayesian approach. *Social Networks*, **25**, 103–140.
- Contisciani, M., Safdari, H. and De Bacco, C. (2021) Community detection and reciprocity in networks by jointly modeling pairs of edges. *arXiv:2112.10436*.
- De Bacco, C., Power, E. A., Larremore, D. B. and Moore, C. (2017) Community detection, link prediction, and layer interdependence in multilayer networks. *Physical Review E*, **95**, 042317.
- D'haeseleer, P. and Church, G. M. (2004) Estimating and improving protein interaction error rates. In *Proceedings*. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004., 216–223. IEEE.
- Eagle, D. E. and Proeschold-Bell, R. J. (2015) Methodological considerations in the use of name generators and interpreters. *Social Networks*, **40**, 75–83.
- Eagle, N., Pentland, A. S. and Lazer, D. (2009) Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, **106**, 15274–15278.
- Fehr, E. and Gächter, S. (1998) Reciprocity and economics: The economic implications of Homo Reciprocans. *European Economic Review*, **42**, 845–859.
- Freeman, L. C. (1992) Filling in the blanks: a theory of cognitive categories and the structure of social affiliation. *Social Psychology Quarterly*, **55**, 118–127.
- Givens, C. R. and Shortt, R. M. (1984) A class of wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, **31**, 231–240.
- Heider, F. (1958) The psychology of interpersonal relations. New York, NY: Wiley.
- Hobson, E. A., Silk, M. J., Fefferman, N. H., Larremore, D. B., Rombach, P., Shai, S. and Pinter-Wollman, N. (2020) A guide to choosing and implementing reference models for social network analysis. arXiv:2012.04720.
- Hoff, P., Fosdick, B., Volfovsky, A. and Stovel, K. (2013) Likelihoods for fixed rank nomination networks. *Network Science*, **1**, 253–277.
- Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983) Stochastic blockmodels: First steps. *Social Networks*, 5, 109–137.
- Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T. and Stephenson, D. (2015) Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10, 227–237.
- 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. and Tan, X. (2012) Social capital and social quilts: network patterns of favor exchange. *American Economic Review*, **102**, 1857–1897.
- Karrer, B. and Newman, M. E. J. (2011) Stochastic blockmodels and community structure in networks. *Physical Review E*, **83**, 016107.
- Kenny, D. A. and La Voie, L. (1984) The social relations model. In *Advances in Experimental Social Psychology*, vol. 18, 141–182. Academic Press.
- Killworth, P. and Bernard, H. (1976) Informant accuracy in social network data. Human Organization, 35, 269–286.
- Koster, J. (2018) Family ties: the multilevel effects of households and kinship on the networks of individuals. *Royal Society Open Science*, **5**, 172159.
- Krackhardt, D. (1987) Cognitive social structures. Social Networks, 9, 109–134.
- Krackhardt, D. and Kilduff, M. (1990) Friendship patterns and culture: the control of organizational diversity. *American Anthropologist*, **92**, 142–154.
- Kristiansen, S. (2004) Social networks and business success: the role of subcultures in an african context. *American Journal of Economics and Sociology*, **63**, 1149–1171.
- Lazega, E. and Snijders, T. A. B. (eds.) (2015) Multilevel network analysis for the social sciences: theory, methods and applications. Oxford: Springer.
- Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K. and Radford, J. (2021) Meaningful measures of human society in the twenty-first century. *Nature*, 595, 189–196.
- Lee, F. and 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.
- 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. and Contractor, N. S. (2021) Using Trellis software to enhance high-quality large-scale network data collection in the field. *Social Networks*, **66**, 171–184.
- Marin, A. (2004) Are respondents more likely to list alters with certain characteristics?: Implications for name generator data. *Social Networks*, **26**, 289–307.
- Marineau, J. E., Labianca, G. J., Brass, D. J., Borgatti, S. P. and 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, 435-463.
- (2003) Interviewer effects in measuring network size using a single name generator. Social Networks, 25, 1–16.
- (2005) Recent developments in network measurement. In *Models and methods in social network analysis* (eds. P. J. Carrington, J. Scott and S. Wasserman), no. 27 in Structural analysis in the social sciences, 8–30. Cambridge; New York: Cambridge University Press.
- Molm, L. D. (2010) The structure of reciprocity. Social Psychology Quarterly, 73, 119–131.
- Newcomb, T. M. (1961) The acquaintance process as a prototype of human interaction. In *The Acquaintance Process* (ed. T. M. Newcomb), 259–261. Holt, Rinehart & Winston.
- Newman, M. E. J. (2018a) Estimating network structure from unreliable measurements. *Physical Review E*, 98, 062321.
- (2018b) Network structure from rich but noisy data. *Nature Physics*, 14, 542–545.
- Nolin, D. A. (2008) Food-Sharing Networks in Lamalera, Indonesia: Tests of Adaptive Hypotheses. Ph.D., University of Washington, Washington, US.

- (2010) Food-sharing networks in Lamalera, Indonesia: Reciprocity, kinship, and distance. *Human Nature*, **21**, 243–268.
- Park, P. S., Blumenstock, J. E. and Macy, M. W. (2018) The strength of long-range ties in population-scale social networks. *Science*, 362, 1410–1413.
- Peixoto, T. P. (2018) Reconstructing networks with unknown and heterogeneous errors. Physical Review X, 8, 041011.
- (2019) Network reconstruction and community detection from dynamics. *Physical Review Letters*, **123**, 128301.
- (2021) Disentangling homophily, community structure and triadic closure in networks. *arXiv preprint arXiv:2101.02510*.
- Perkins, J. M., Subramanian, S. V. and 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.
- Power, E. A. (2017) Social support networks and religiosity in rural South India. Nature Human Behaviour, 1, 0057.
- Pustejovsky, J. E. and Spillane, J. P. (2009) Question-order effects in social network name generators. *Social Networks*, **31**, 221–229.
- Ready, E. and Power, E. A. (2018) Why wage earners hunt: Food sharing, social structure, and influence in an Arctic mixed economy. *Current Anthropology*, **59**, 74–97.
- (2021) Measuring reciprocity: Double sampling, concordance, and network construction. Network Science, 1-16.
- Redhead, D., McElreath, R. and Ross, C. T. (2021) Reliable network inference from unreliable data: A tutorial on latent network modeling using STRAND. *PsyArXiv*. URL: https://psyarxiv.com/mkp2y/.
- Redhead, D. and Power, E. A. (Accepted) Social hierarchies and social networks in humans. *Philosophical Transaction* of the Royal Society B.
- Redhead, D. and von Rueden, C. R. (2021) Coalitions and conflict: A longitudinal analysis of men's politics. *Evolutionary Human Sciences*, **3**.
- Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007) An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29, 173–191.
- Ross, C. T. and Redhead, D. (2021) DieTryin: An R package for data collection, automated data entry, and postprocessing of network-structured economic games, social networks, and other roster-based dyadic data. *Behavior Research Methods*.
- von Rueden, C., Redhead, D., O'Gorman, R., Kaplan, H. and 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**, 20191367.
- Safdari, H., Contisciani, M. and De Bacco, C. (2021a) Generative model for reciprocity and community detection in networks. *Phys. Rev. Research*, **3**, 023209.
- Safdari, H., Contisciani, M. and De Bacco, C. (2021b) Reciprocity, community detection, and link prediction in dynamic networks. *arXiv:2112.09624*.
- Shakya, H. B., Christakis, N. A. and Fowler, J. H. (2017) An exploratory comparison of name generator content: Data from rural India. *Social Networks*, **48**, 157–168.
- Simpson, B., Markovsky, B. and Steketee, M. (2011) Power and the perception of social networks. *Social Networks*, **33**, 166–171.
- Simpson, C. R. (2020) On the structural equivalence of coresidents and the measurement of village social structure. *Social Networks*.

Smith, K. P. and Christakis, N. A. (2008) Social networks and health. Annual Review of Sociology, 34, 405–429.

- Snijders, T. A. and Kenny, D. A. (1999) The social relations model for family data: A multilevel approach. *Personal Relationships*, **6**, 471–486.
- Sprinzak, E., Sattath, S. and Margalit, H. (2003) How reliable are experimental protein–protein interaction data? *Journal of Molecular Biology*, **327**, 919–923.
- Titterington, D., Smith, A. and Makov, U. (1985) Statistical Analysis of Finite Mixture Distributions. Wiley, New York.
- Warner, R. M., Kenny, D. A. and Stoto, M. (1979) A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, **37**, 1742–1757.
- Young, J.-G., Cantwell, G. T. and Newman, M. E. J. (2021a) Bayesian inference of network structure from unreliable data. *Journal of Complex Networks*, **8**, cnaa046.
- Young, J.-G., Kirkley, A. and Newman, M. (2021b) Clustering of heterogeneous populations of networks. *arXiv* preprint arXiv:2107.07489.
- Young, J.-G., Valdovinos, F. S. and Newman, M. (2021c) Reconstruction of plant–pollinator networks from observational data. *Nature Communications*, **12**, 1–12.
- Yousefi-Nooraie, R., Marin, A., Hanneman, R., Pullenayegum, E., Lohfeld, L. and Dobbins, M. (2019) The relationship between the position of name generator questions and responsiveness in multiple name generator surveys. *Sociological Methods & Research*, 48, 243–262.

Supporting Information (SI)

S1.1 Variational Inference

Variational Inference assumes approximate posterior distributions for the parameters (Blei et al., 2017). We assume the mean-field variational family:

$$q(\eta, \lambda, \theta, Y) = q(\eta; \nu^{shape}, \nu^{rate}) \prod_{k,} q(\lambda_k; \phi_k^{shape}, \phi_k^{rate}) \prod_{m,} q(\theta_m; \gamma_m^{shape}, \gamma_m^{rate}) \prod_{i,j,} q(Y_{ij}; \rho_{ij}),$$
(S2)

where $q(\eta; \nu^{shape}, \nu^{rate})$, $q(\lambda_k; \phi_k^{shape}, \phi_k^{rate})$ and $q(\theta_m; \gamma_m^{shape}, \gamma_m^{rate})$ are Gamma distributions, and $q(Y_{ij}; \rho_{ij})$ is a Categorical distribution. The idea is to set the parameters of these distribution such that the KL divergence between this approximate posterior and the true posterior is minimized. This is equivalent to maximize the Evidence Lower Bound (ELBO), defined as:

$$\mathsf{ELBO}(q) = \mathbb{E}_q \left[\log \mathcal{L}(\lambda, \theta, \eta, Y) \right] - \mathbb{E}_q \left[\log q(\eta, \lambda, \theta, Y) \right] \,. \tag{S3}$$

This could be done by taking the derivative w.r.t. each parameter of the variational distributions individually and setting them to zero. However, there is an alternative more direct way to extract the optimal parameters, which involves finding the complete conditionals of each variable (Blei et al., 2017). This is defined as the conditional distribution of a given parameter, given all the remaining parameters and data.

S1.1.1 Auxiliary variables

To facilitate inference we augment the model by introducing auxiliary variables that allow to make the model conditionally conjugate, i.e. all the complete conditionals are in the exponential family, which will eventually allow close-form and faster parameters' updates. We define discrete latent auxiliary variables:

$$z_{mk}^1 \sim \operatorname{Pois}(\theta_m \lambda_k), \quad z_{ijm}^2 \sim \operatorname{Pois}(\eta X_{jim}) \quad s.t. \quad z_{mk}^1 + z_{ijm}^2 = X_{ijm} \quad | \quad Y_{ij} = k \;. \tag{S4}$$

With these new variables one can work with an augmented conditional distribution $P(X_{ijm}, Z|X_{jim}, Y_{ij} = k, \lambda_k, \theta_m, \eta)$ which is factorized and thus easier to perform inference with. Notably, this joint conditional distribution is built such that its marginal recovers the original conditional distribution:

$$P(X_{ijm}|X_{jim}, Y_{ij} = k, \lambda_k, \theta_m, \eta) = \sum_{\substack{z_{mk}^1, z_{ijm}^2 \\ z_{im}^1, z_{ijm}}} P(X_{ijm}, Z|X_{jim}, Y_{ij} = k, \lambda_k, \theta_m, \eta) .$$
(S5)

This comes from the additive property of Poisson distributions: the sum of i.i.d. Poisson distributed variables is also Poisson distributed, with parameter the sum of the individual variables' parameters. In fact:

$$P(X_{ijm}, Z|X_{jim}, Y_{ij} = k, \lambda_k, \theta_m, \eta) = P(X_{ijm}|Z, Y_{ij} = k) P(Z|X_{jim}, \lambda_k, \theta_m, \eta)$$
(S6)

$$P(X_{ijm}|Z, Y_{ij} = k) = \delta \left(z_{mk}^1 + z_{ijm}^2 - X_{ijm} \right)^{Y_{ij,k}}$$
(S7)

$$P(Z|X_{jim}, \lambda_k, \theta_m, \eta) = \operatorname{Pois}(z_{mk}^1; \theta_m \lambda_k) \operatorname{Pois}(z_{ijm}^2; \eta X_{jim}), \qquad (S8)$$

where $\delta(\cdot)$ is the Dirac delta distribution. The variational distribution for Z is a Multinomial distribution

$$q\left((z_{mkl}^{1}, z_{ijml}^{2}); X_{ijm}, \left[\hat{z}_{mkl}^{1}, \hat{z}_{ijml}^{2}\right]\right) .$$
(S9)

Assuming $z_{mkl}^1, z_{ijml}^2, \eta, \lambda_k, \theta_m$ and Y_{ij} are all independent, our goal is to infer the parameters of their variational distributions. When all the complete conditionals are in the exponential family, we can use the result of Blei et al. (2017) that the natural parameters ω_i of the variational family satisfy:

$$\omega_i = \mathbb{E}_{q(y)} \kappa_i(y) , \qquad (S10)$$

where y is the parameter from the original posterior and $\kappa_i(y)$ is the natural parameter of the complete conditional. The expectation is with respect to the variational distribution q(y). The natural parameters for a Gamma distribution $\text{Gam}(\alpha, \beta)$ are $(\alpha - 1, -\beta)$; for a Multinomial $\text{Mult}(n, [\log p_1, \ldots, \log p_K])$ and a Categorical distribution $\text{Cat}([p_1, \ldots, p_K])$ are $(\log p_1, \ldots, \log p_K)$. The full posterior distribution of the augmented system is:

$$P(Y, \lambda, \theta, \eta | X, Z) = \prod_{i,j,,m,k} \left[\left(\text{Pois}(z_{mk}^{1}; \theta_{m} \lambda_{k}) \text{Pois}(z_{ijm}^{2}; \eta X_{jim}) \right)^{Y_{ij,k}} \delta \left(z_{mk}^{1} + z_{ijm}^{2} - X_{ijm} \right)^{Y_{ij,k}} \right] \times \underbrace{\text{Gam}(\eta; c, d) \prod_{k, \text{Gam}(\lambda_{k}; a_{k}, b_{k}) \prod_{m, \text{Gam}(\theta_{m}; \alpha_{m}, \beta_{m})} \prod_{i,j, \text{Cat}(Y_{ij}; p_{ij})} .$$

$$\underbrace{\text{Friors}}$$

$$(S11)$$

S1.1.2 Update of Z

The complete conditional for Z is:

$$P(z_{mk}^{1}, z_{ijm}^{2} | Z_{\backslash z_{mk}^{1}, z_{ijm}^{2}}, X, Y, \lambda, \theta, \eta)$$

$$\propto \delta \left(z_{mk}^{1} + z_{ijm}^{2} - X_{ijm} \right)^{Y_{ij,k}} \left[\frac{X_{ijm}!}{z_{mk}^{1}! z_{ijm}^{2}!} (\theta_{m} \lambda_{k})^{z_{mk}^{1}} (\eta X_{jim})^{z_{ijm}^{2}} \right]^{Y_{ij,k}}$$
(S12)

~
$$\left[\operatorname{Mult}(X_{ijm}, \bar{z}_{ijm})\right]^{Y_{ij,k}}$$
, (S13)

where $\bar{z}_{ijm} = (\theta_m \lambda_k, \eta X_{jim})$ are the parameters of the complete conditional. Using Eq. (S10) we get:

$$\log \hat{z}_{mk}^{1} = \mathbb{E}_{q(\theta_{m})} \left[\log \theta_{m} \right] + \mathbb{E}_{q(\lambda_{k})} \left[\log \lambda_{k} \right]$$

$$= \Psi(\gamma_{m}^{shape}) - \log \gamma_{m}^{rate} + \Psi(\phi_{k}^{shape}) - \log \phi_{k}^{rate} , \qquad (S14)$$

where $\Psi(x)$ is the di-gamma function. Here we used $\mathbb{E}[\log x] = \Psi(a) - \log(b)$, valid for gamma-distributed variables. Similarly:

$$\log \hat{z}_{ijm}^2 = \mathbb{E}_{q(\eta)} \left[\log \eta X_{jim} \right] = \Psi(\nu^{shape}) - \log \nu^{rate} + \log X_{jim} .$$
(S15)

S1.1.3 Update of mutuality parameter η

The complete conditional for η is:

$$P(\eta|X, Z, Y, \lambda, \theta) \propto \operatorname{Gam}(\eta; c, d) \prod_{i,j,,m,k} \operatorname{Pois}(z_{ijm}^2; \eta X_{jim})^{Y_{ij,k}}$$
$$\propto \eta^{c-1} e^{-d\eta} \prod_{i,j,,m,k} \left[(\eta X_{jim})^{z_{ijm}^2} e^{-\eta X_{jim}} \right]^{Y_{ij,k}}$$
(S16)

~
$$\operatorname{Gam}\left(c + \sum_{i,j,m,k} Y_{ij,k} z_{ijm}^{2}, d + \sum_{i,j,m,k} Y_{ij,k} X_{jim}\right)$$
. (S17)

Using Eq. (S10) we get:

$$\nu^{shape} = c + \sum_{i,j,m,k} \rho_{ij,k} X_{ijm} \hat{z}_{ijm}^2$$
(S18)

$$\nu^{rate} = d + \sum_{i,j,m,k} \rho_{ij,k} X_{jim}.$$
(S19)

S1.1.4 Update of λ_k

The complete conditional for λ_k is:

$$P(\lambda_{k}|\lambda_{\lambda_{k}}, X, Z, Y, \theta, \eta) \propto \operatorname{Gam}(\lambda_{k}; a_{k}, b_{k}) \prod_{i,j,m} \operatorname{Pois}(z_{mk}^{1}; \theta_{m}\lambda_{k})^{Y_{ij,k}} \\ \propto (\lambda_{k})^{a_{k}-1} e^{-b_{k}\lambda_{k}} \prod_{i,j,m} \left[(\theta_{m}\lambda_{k})^{z_{mk}^{1}} e^{-\theta_{m}\lambda_{k}} \right]^{Y_{ij,k}}$$

$$(S20)$$

~
$$\operatorname{Gam}\left(a_k + \sum_{i,j,m} Y_{ij,k} \, z_{mk}^1, b_k + \sum_{i,j,m} Y_{ij,k} \, \theta_m\right)$$
 (S21)

Using Eq. (S10) we get:

$$\phi_k^{shape} = a_k + \sum_{i,j,m} \rho_{ij,k} X_{ijm} \hat{z}_{mk}^1$$
(S22)

$$\phi_k^{rate} = b_k + \sum_{i,j,m} \rho_{ij,k} \frac{\gamma_m^{shape}}{\gamma_m^{rate}} \,. \tag{S23}$$

S1.1.5 Update of reliabilities θ_m

The complete conditional for θ_m is:

$$P(\theta_{m}|\theta_{\backslash\theta_{m}}, X, Z, Y, \lambda, \eta) \propto \operatorname{Gam}(\theta_{m}; \alpha_{m}, \beta_{m}) \prod_{i,j,k} \operatorname{Pois}(z_{mk}^{1}; \theta_{m}\lambda_{k})^{Y_{ij,k}} \\ \propto (\theta_{m})^{\alpha_{m}-1} e^{-\beta_{m}} \theta_{m} \prod_{i,j,k} \left[(\theta_{m}\lambda_{k})^{z_{mk}^{1}} e^{-\theta_{m}\lambda_{k}} \right]^{Y_{ij,k}}$$
(S24)

~
$$\operatorname{Gam}\left(\alpha_m + \sum_{i,j,k} Y_{ij,k} z_{mk}^1, \beta_m + \sum_{i,j,k} Y_{ij,k} \lambda_k\right)$$
 (S25)

Using Eq. (S10) we get:

$$\gamma_m^{shape} = \alpha_m + \sum_{i,j,k} \rho_{ij,k} X_{ijm} \hat{z}_{mk}^1$$
(S26)

$$\gamma_m^{rate} = \beta_m + \sum_{i,j,k} \rho_{ij,k} \frac{\phi_k^{shape}}{\phi_k^{rate}} \,. \tag{S27}$$

S1.1.6 Update of true ties Y_{ij}

We have a complete conditional:

$$P(Y_{ij}|Y_{\backslash Y_{ij}}, X, Z, \lambda, \theta, \eta) \propto$$

$$\propto P(Y_{ij}) \prod_{k} \left[\prod_{m} \left(\text{Pois}(z_{mk}^{1}; \theta_{m}\lambda_{k}) \operatorname{Pois}(z_{ijm}^{2}; \eta X_{jim}) \right)^{Y_{ij,k}} \delta\left(z_{mk}^{1} + z_{ijm}^{2} - X_{ijm} \right)^{Y_{ij,k}} \right]$$

$$\propto \prod_{k} \left[p_{ij,k} \prod_{m} \left(\operatorname{Pois}(z_{mk}^{1}; \theta_{m}\lambda_{k}) \operatorname{Pois}(z_{ijm}^{2}; \eta X_{jim}) \right) \delta\left(z_{mk}^{1} + z_{ijm}^{2} - X_{ijm} \right) \right]^{Y_{ij,k}}$$

$$\sim \operatorname{Cat}\left(Y_{ij}; \rho_{ij} \right)$$
(S28)

LATENT NETWORK MODELS TO ACCOUNT FOR NOISY, MULTIPLY-REPORTED SOCIAL NETWORK DATA

$$\rho_{ij,k} = p_{ij,k} \prod_{m} \operatorname{Pois}(z_{mk}^1; \theta_m \lambda_k) \operatorname{Pois}(z_{ijm}^2; \eta X_{jim}) \delta\left(z_{mk}^1 + z_{ijm}^2 - X_{ijm}\right) .$$
(S30)

Using Eq. (S10) we get:

 $\rho_{ij,k}$

$$\propto \exp \left\{ \log p_{ij,k} + \sum_{m} \left(X_{ijm} \hat{z}_{mk}^{1} \left(\mathbb{E}_{q(\theta_{m})} \left[\log \theta_{m} \right] + \mathbb{E}_{q(\lambda_{k})} \left[\log \lambda_{k} \right] \right) \right) + \right. \\ \left. + \sum_{m} X_{ijm} \hat{z}_{ijm}^{2} \mathbb{E}_{q(\eta)} \left[\log \eta X_{jim} \right] + \right. \\ \left. - \frac{\phi_{k}^{shape}}{\phi_{k}^{rate}} \sum_{m} \frac{\gamma_{m}^{shape}}{\gamma_{m}^{rate}} - \frac{\nu^{shape}}{\nu^{rate}} \sum_{m} X_{jim} + \right. \\ \left. - \sum_{m} \mathbb{E}_{q(z_{mk}^{1})} \left[\log z_{mk}^{1} \right] - \sum_{m} \mathbb{E}_{q(z_{ijm}^{2})} \left[\log z_{ijm}^{2} \right] \right\} .$$
(S31)

Given that we have to normalize such that $\sum_k \rho_{ij,k} = 1$, we can shorten the equation by considering only terms that depend on k:

$$\rho_{ij,k} \propto \exp\left\{\log p_{ij,k} + \sum_{m} \left(X_{ijm} \hat{z}_{mk}^{1} \mathbb{E}_{q(\lambda_k)} \left[\log \lambda_k\right]\right) - \frac{\phi_k^{shape}}{\phi_k^{rate}} \sum_{m} \frac{\gamma_m^{shape}}{\gamma_m^{rate}}\right\},$$
(S32)

where we used $z_{mk}^1 = X_{ijm} - z_{ijm}^2$ to remove terms that depend on k only through z_{mk}^1 .

S1.1.7 ELBO

The Evidence Lower Bound (ELBO) is defined as:

$$\mathsf{ELBO}(q) = \mathbb{E}_q \left[\log \mathcal{L}(\lambda, \theta, \eta, Y) \right] - \mathbb{E}_q \left[\log q(\eta, \lambda, \theta, Y) \right] \,, \tag{S33}$$

where $\mathcal{L}(\lambda, \theta, \eta, Y)$ and $q(\eta, \lambda, \theta, Y)$ are given by Eq. (5) and Eq. (82). Thus,

 $\text{ELBO}(q) \propto$

$$\mathbb{E}_{q} \left[\sum_{i,j,l,m,k} Y_{ij,k} \left(X_{ijm} \log(\theta_{m} \lambda_{k} + \eta X_{jim}) - (\theta_{m} \lambda_{k} + \eta X_{jim}) \right) \right] + \\ + \mathbb{E}_{q} \left[\sum_{i,j,l,k} Y_{ij,k} \log(p_{ij,k}) \right] + \\ + \mathbb{E}_{q} \left[c \log(\eta) - d\eta + \sum_{k,l} (a_{kl} \log(\lambda_{k}) - b_{kl} \lambda_{k}) + \sum_{m,l} (\alpha_{ml} \log(\theta_{m}) - \beta_{ml} \theta_{m}) \right] + \\ - \mathbb{E}_{q} \left[\sum_{i,j,l,k} \left(Y_{ij,k} \log(\rho_{ij,k}) \right) + \nu^{shape} \log(\nu^{rate}) - \log(\Gamma(\nu^{shape})) + \nu^{shape} \log(\eta) - \nu^{rate} \eta \right] + \\ - \mathbb{E}_{q} \left[\sum_{k,l} \left(\phi_{k}^{shape} \log(\phi_{k}^{rate}) - \log(\Gamma(\phi_{k}^{shape})) + \phi_{k}^{shape} \log(\lambda_{k}) - \phi_{k}^{rate} \lambda_{k} \right) \right] + \\ - \mathbb{E}_{q} \left[\sum_{n,l} \left(\gamma_{m}^{shape} \log(\gamma_{m}^{rate}) - \log(\Gamma(\gamma_{m}^{shape})) + \gamma_{m}^{shape} \log(\theta_{m}) - \gamma_{m}^{rate} \theta_{m} \right) \right],$$
(S34)

where the proportionality is because we are omitting constant terms. Remember:

$$\mathbb{E}_{q}\left[Y_{ij,k}\right] = \rho_{ij,k}$$

$$\gamma^{shape}$$
(S35)

$$\mathbb{E}_{q}\left[\theta_{m}\right] = \frac{\gamma_{m}}{\gamma_{m}^{rate}} \tag{S36}$$

$$\mathbb{E}_{q}\left[\lambda_{k}\right] = \frac{\phi_{k}^{shape}}{\phi_{k}^{rate}}$$
(S37)

$$\mathbb{E}_{q}\left[\eta\right] = \frac{\nu^{shape}}{\nu^{rate}} \tag{S38}$$

$$\mathbb{E}_{q}\left[\log(\theta_{m})\right] = \Psi(\gamma_{m}^{shape}) - \log(\gamma_{m}^{rate})$$
(S39)

$$\mathbb{E}_{q}\left[\log(\lambda_{k})\right] = \Psi(\phi_{k}^{shape}) - \log(\phi_{k}^{rate}) \tag{S40}$$

$$\mathbb{E}_q \left[\log(\eta) \right] = \Psi(\nu^{snape}) - \log(\nu^{rate}) .$$
(S41)

Moreover, we use the approximation:

$$\mathbb{E}_{q}\left[\log(\theta_{m}\lambda_{k}+\eta X_{jim})\right] = \mathbb{E}_{q}\left[\log(\theta_{m}\lambda_{k})\right] + \mathbb{E}_{q}\left[\log(\eta X_{jim})\right] , \qquad (S42)$$

$$\mathbb{E}_{q}\left[\log(\theta_{m}\lambda_{k}+\eta X_{jim})\right] \text{ and } \mathbb{E}_{q}\left[\log(\theta_{m}\lambda_{k})\right] + \mathbb{E}_{q}\left[\log(\eta X_{jim})\right] + \mathbb{E}_{q}\left[\log(\eta X_{jim})\right] , \qquad (S42)$$

where
$$\mathbb{E}_{q} \left[\log(\theta_{m} \lambda_{k}) \right]$$
 and $\mathbb{E}_{q} \left[\log(\eta X_{jim}) \right]$ are defined in Eq. (S14) and $\operatorname{Eq.}_{pq}(S15)$. Then, the ELBO is given by:

$$\sum_{\substack{j,j,l,m,k \\ q}} \sum_{\substack{j,j,l,m,k \\ q}} \left[\lambda_{ijm} \left(\mathbb{E}_{q} \left[\log(\theta_{m} \lambda_{k}) \right] + \mathbb{E}_{q} \left[\log(\eta X_{jim}) \right] \right) - \left(\frac{\gamma_{m}}{\gamma_{m}^{rate}} \frac{\phi_{k}}{\phi_{k}^{rate}} + \frac{\nu}{\nu^{rate}} X_{jim} \right) \right) \right] + \frac{1}{2} \left[\operatorname{ELBO}(q)^{i} \left[X_{jim} \left(\mathbb{E}_{q} \left[\log(\theta_{m} \lambda_{k}) \right] + \mathbb{E}_{q} \left[\log(\eta X_{jim}) \right] \right) - \left(\log(\eta X_{jim}) \right] \right] - \left(\frac{\gamma_{m}}{\gamma_{m}^{rate}} \frac{\phi_{k}}{\phi_{k}^{rate}} + \frac{\nu}{\nu^{rate}} X_{jim} \right) \right) \right] + \frac{1}{2} \left[\operatorname{ELBO}(q)^{i} \left[X_{jim} \left(\mathbb{E}_{q} \left[\log(\theta_{m} \lambda_{k}) \right] + \log(\Gamma(\nu^{shape})) - c \log(\nu^{rate}) + \nu^{shape} \left(1 - \frac{d}{\nu^{rate}} \right) + \frac{1}{2} \left[\Psi(\phi_{k}^{shape})(a_{kl} - \phi_{k}^{shape}) + \log(\Gamma(\phi_{k}^{shape})) - a_{kl} \log(\phi_{k}^{rate}) + \phi_{k}^{shape} \left(1 - \frac{b_{kl}}{\phi_{k}^{rate}} \right) \right] \right] + \frac{1}{2} \sum_{k,l} \left[\Psi(\gamma_{m}^{shape})(\alpha_{ml} - \gamma_{m}^{shape}) + \log(\Gamma(\gamma_{m}^{shape})) - \alpha_{ml} \log(\gamma_{m}^{rate}) + \gamma_{m}^{shape} \left(1 - \frac{\beta_{ml}}{\phi_{k}^{rate}} \right) \right] \right] + \frac{1}{2} \sum_{i,j,l,k} \left[\rho_{ij,k} \left(\log(p_{ij,k}) - \log(\rho_{ij,k}) \right) \right] \right].$$
(S43)

S1.2 Deriving the expected value of the marginal distribution

$$\mathbb{E}\left[X_{ijm}|Y_{ij}=k\right] = \mu_{ijm} = \sum_{X_{ijm}, X_{jim}} X_{ijm} P(X_{ijm}, X_{jim}|\Theta)$$

$$= \sum_{X_{jim}} P(X_{jim}|\Theta) \sum_{X_{ijm}} X_{ijm} P(X_{ijm}|X_{jim},\Theta)$$

$$= \sum_{X_{jim}} P(X_{jim}|\Theta) \left[\theta_m \lambda_{k|Y_{ij}=k} + \eta X_{jim}\right]$$

$$= \theta_m \lambda_{k|Y_{ij}=k} + \eta \sum_{X_{jim}} X_{jim} P(X_{jim}|\Theta)$$

$$= \theta_m \lambda_{k|Y_{ij}=k} + \eta \mu_{jim}$$

$$= \theta_m \lambda_{k|Y_{ij}=k} + \eta \left(\theta_m \lambda_{k|Y_{ji}=k} + \eta \mu_{ijm}\right) \quad .$$
(S44)

Solving for μ_{ijm} yields:

$$\mu_{ijm} \left(1 - \eta^2 \right) = \left(\theta_m \lambda_{k|Y_{ij}=k} + \eta \, \theta_m \lambda_{k|Y_{ji}=k} \right) \quad , \tag{S45}$$

which implies:

$$\mu_{ijm} = \frac{\theta_m \lambda_{k|Y_{ij}=k} + \eta \,\theta_m \lambda_{k|Y_{ji}=k}}{(1-\eta^2)} \quad .$$
(S46)

S1.3 Synthetic data generation

The synthetic networks used in the analysis are of three types and represent different scenarios: networks with fixed reliability and a fraction of over-reporters (a), networks with fixed reliability and a fraction of under-reporters (b), and networks with randomly generated reliabilities θ_m (c). In these experiments, the true unobserved directed and binary networks Y are generated by using either a stochastic block model (a), (b), or a degree-corrected stochastic block model (c). In all cases, the networks have N = 100 nodes, C = 2 non-overlapping communities of equal size, and average degree $\langle k \rangle = 10$. In the scenario (c), the degree sequences are extracted from a power law distribution with exponents equal to 2 and 2.5 for the in-degree and the out-degree respectively. The networks follow an assortative structure and tie probabilities are $p_{in} = \langle k \rangle C/N$ and $p_{out} = 0.1 p_{in}$, for within-group and between-group ties. To generate the observed directed and weighted double-sampled networks X, we need additional parameters to encode the individual reliability θ , the mutuality effect η and the contribution of the underlying data Y in determining X, i.e. λ . Then, the networks X are generated in a two-step routine:

- 1. Select with a coin-flip one direction, (i, j) or (j, i). Say we select (i, j).
- 2. Sample X_{iim} from the marginal $P(X_{iim}|\Theta) = \text{Pois}(\mu_{iim})$, where

$$\mu_{ijm} = \frac{\theta_m \lambda_{k|Y_{ij}=k} + \eta \, \theta_m \lambda_{k|Y_{ji}=k}}{1 - \eta^2}$$

is the mean of the marginal such that it is consistent with the joint and the conditional distributions. For further details see Section S1.2.

3. Sample X_{jim} from the conditional $P(X_{jim}|X_{ijm}, Y_{ji} = k, \lambda_k, \theta_m, \eta) = \text{Pois}(\theta_m \lambda_k + \eta X_{ijm})$ using the previously extracted value of X_{iim} .

We use M = 100 reporters and a reporter m only reports ties involving her, i.e. $R_{ijm} = 0$ when $m \notin \{i, j\}$. In this equation, X is weighted while Y is binary so $\lambda = [\lambda_0, \lambda_1]$. Furthermore, in these experiments we consider single-layer networks, i.e. L = 1.

In scenarios (a) and (b) we vary the fraction of non-reliable individuals $\theta_{ratio} \in [0, 0.1, \dots, 0.5]$, where we avoid values bigger than 0.5 because they would result in unlikely and non-meaningful regimes. The reliability for this fraction of people is $\theta_m = 50$ and $\theta_m = 0.5$ for the over- and under-reporters respectively. All the other subjects have $\theta_m = 1$ and entries of X determined deterministically by using Eq. (1). In these settings $\lambda = [0.01, 1.0]$, while $\eta \in \{0, 0.2\}$ and $\eta \in \{0, 0.6\}$ in scenarios (a) and (b) respectively. The different values of the mutuality parameter are used to provide comparable behaviours in the two regimes that are different because of the magnitudes of the mean of the Poisson.

In the third scenario, θ_m is sampled from a Gamma distribution with $\alpha = \beta = 2$, the mutuality effect is given and varies as $\eta \in \{0, 0.2, 0.6\}$, while $\lambda_0 = 0.01$ and $\lambda_1 = \lambda_0 + \lambda_{diff}$ with $\lambda_{diff} \in [0.1, 0.2, \dots, 1]$.

S1.4 Recoverability of latent network properties in synthetic experiments

S1.4.1 Planted mutuality vs inferred mutuality

From the experiments with synthetic networks, we observe that VIMuRe_T is capable at distinguishing between the scenarios of low and high mutuality on respondents' data, X. In fact, the mutuality inferred by our model (η_{est}) is highly correlated with the planted ground truth mutuality (η_{GT}), as it can be seen in Figure S3. Mutuality inferred by our model is therefore reliable, requiring at most a downward adjustment when interpreting network characteristics.



Figure S2: Synthetic networks with over- or under-reporters: estimating reliabilities θ . Synthetic networks with N = 100 nodes and M = 100 reporters generated with the benchmark generative model proposed above 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 over ten independent synthetic networks, and the accuracy over the reliabilities θ is measured with the Mean Squared Error. Values close to 0 indicate perfect recovery.



Figure S3: Synthetic networks with N = 100 nodes and M = 100 reporters generated with the benchmark generative model with $\lambda_1 - \lambda_0 = 1.0$ and planted reciprocity values around ≈ 0.2 on the ground truth networks. The plot compares the distributions of mutuality values as recovered by VIMuRe_T (η_{est}) to the true planted mutuality on X, η_{GT} .



Figure S4: Synthetic networks with N = 100 nodes and M = 100 reporters generated with the benchmark generative model with $\lambda_1 - \lambda_0 = 1.0$ and planted reciprocity values around ≈ 0.2 on the ground truth network. The four sub-plots represent networks generated with low (top left), medium (top right) to increasingly high (bottom left and right) mutuality effect η . The box plots are distributions of the network density on the Y, over twenty samples of synthetic networks.



Figure S5: Distribution of reliability estimates across different tie types in Karnataka networks.



Figure S6: Wasserstein distance of reporters' reliabilities between different tie types in Karnataka networks, measured by the Wasserstein metric.



Figure S7: In-degree (receive) vs individual reliability in Karnataka networks per tie type.



Figure S8: Distribution of reliability estimates of respondent nodes in the Nicaragua dataset, obtained after the 2-step process.

LATENT NETWORK MODELS TO ACCOUNT FOR NOISY, MULTIPLY-REPORTED SOCIAL NETWORK DATA

Networks	Edges			Mean Degree			Transitivity			Reciprocity			
	(mean	\pm	std)	(mean	\pm	std)	(mean	\pm	std)	(mean	\pm	std)	
Karnataka													
(75 villages)													
Advice receive	243.04	±	83.92	1.21	\pm	0.16	0.14	±	0.06	0.20	±	0.08	
Advice give	195.96	\pm	71.39	1.05	\pm	0.11	0.04	\pm	0.02	0.22	\pm	0.09	
Advice (intersection)	44.69	\pm	23.51	0.20	\pm	0.09	0.06	\pm	0.16	0.45	\pm	0.18	
Advice (union)	394.31	\pm	137.06	1.74	\pm	0.26	0.11	\pm	0.04	0.50	\pm	0.11	
VIMuRe	343.57	±	117.39	1.52	±	0.21	0.11	±	0.03	0.39	±	0.07	
Money receive	249.03	±	102.08	1.28	\pm	0.20	0.13	±	0.06	0.15	±	0.05	
Money give	217.53	\pm	88.44	1.18	\pm	0.17	0.05	\pm	0.03	0.16	\pm	0.05	
Money (intersection)	37.43	\pm	19.56	0.17	\pm	0.07	0.06	\pm	0.14	0.58	\pm	0.17	
Money (union)	429.13	\pm	174.32	1.88	\pm	0.39	0.11	\pm	0.05	0.64	\pm	0.09	
VIMuRe	322.17	±	135.20	1.40	±	0.29	0.09	±	0.04	0.43	±	0.08	
Visit go	344.25	±	142.18	1.64	\pm	0.30	0.16	±	0.06	0.19	±	0.05	
Visit come	328.99	\pm	133.76	1.58	\pm	0.27	0.07	\pm	0.03	0.19	\pm	0.05	
Visit (intersection)	63.97	\pm	31.70	0.28	\pm	0.11	0.14	\pm	0.17	0.70	\pm	0.16	
Visit (union)	609.27	\pm	249.48	2.66	\pm	0.59	0.13	\pm	0.05	0.76	\pm	0.10	
VIMuRe	389.60	±	189.50	1.69	±	0.49	0.11	±	0.04	0.47	±	0.07	
Household items receive	289.08	±	118.49	1.45	±	0.24	0.19	±	0.07	0.20	±	0.05	
Household items give	281.43	\pm	111.24	1.41	\pm	0.22	0.07	\pm	0.03	0.18	\pm	0.05	
Household items (intersection)	54.67	\pm	29.97	0.24	\pm	0.11	0.19	\pm	0.22	0.72	\pm	0.16	
Household items (union)	515.84	\pm	204.45	2.26	\pm	0.46	0.14	\pm	0.05	0.74	\pm	0.12	
VIMuRe	350.68	±	159.07	1.53	±	0.38	0.11	±	0.05	0.48	±	0.08	
Nicaragua													
(108 nodes, 106 respondents)													
Incoming	2595			24.03	±	12.12	0.40			0.46			
Outgoing		2958		27.39	\pm	16.34		0.40			0.55		
Intersection		1422		13.17	\pm	7.69		0.38			0.71		
Union		4131		38.25	\pm	16.82		0.52			0.76		
VIMuRe		1517		14.05	\pm	7.18		0.16			0.11		

Table S2: Summary statistics of original network layers, baseline aggregations and networks inferred by VIMuRe.