Available online at www.sciencedirect.com
SCIENCE DIRECT®
Social Networks 25 (2003) 103–140
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Network inference, error, and informant (in)accuracy: a Bayesian approach☆

Carter T. Butts*,[1]

Department of Sociology, Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA 92697, USA

Abstract

Much, if not most, social network data is derived from informant reports; past research, however, has indicated that such reports are in fact highly inaccurate representations of social interaction. In this paper, a family of hierarchical Bayesian models is developed which allows for the simultaneous inference of informant accuracy and social structure in the presence of measurement error and missing data. Posterior simulation for these models using Markov Chain Monte Carlo methods is outlined. Robustness of the models to structurally correlated error rates, implications of the Bayesian modeling framework for improved data collection strategies, and the validity of the criterion graph are also discussed.

JEL classification: C110 (Mathematical and Quantitative Methods: Bayesian Analysis)

Keywords: Informant accuracy; Measurement error; Hierarchical Bayesian models; Network inference; Data collection strategies

1. Introduction

Despite the considerable success of modern network analysis, a number of basic foundational issues remain to be addressed. Especially problematic is the lack of a systematic treatment of error in network data. Though it is widely understood among network analysts

☆ This material is based upon work supported under a National Science Foundation Graduate Fellowship, and was supported in part by the Center for the Computational Analysis of Social and Organizational Systems and the Institute for Complex Engineered Systems at Carnegie Mellon University.

* Tel.: +1-949-824-8591.
E-mail address: buttsc@uci.edu (C.T. Butts).

[1] This work has benefited from helpful commentary by a number of people, including David Krackhardt, Mark Handcock, Martina Morris, Kathleen Carley, Kim Romney, Bill Batchelder, Matt Dombroski, Benoit Morel, David Rode, Isa Verdinelli, and Peter Marsden. Any weaknesses which remain are the fault of the author.

0378-8733/03/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
PII: S0378-8733(02)00038-2

both that network data is subject to non-trivial levels of error and that network analytic techniques are generally quite sensitive to such error, there is currently no general, agreed-upon framework for dealing with this problem. In this paper, we shall demonstrate a fairly general approach to modeling certain kinds of error which can be used generically to assess posterior uncertainty in any classical network measure or comparison. This approach can ultimately be extended to account for other sources of error (by modifying the error model), and can be utilized to facilitate network sampling, robustness testing, and analysis of incomplete data sets. It is hoped that this work will serve as the basis for further growth of a generic approach to error modeling in social network analysis.

1.1. Self-report data and informant accuracy

The ultimate foundation of any scientific endeavor is the data from which inferences are to be drawn. In the field of social network analysis, the quality of this foundation has long been in question. A large literature on informant accuracy in network analysis—most notably that written by and in response to the work of Bernard, Killworth, and Sailer (BKS)—has debated the question of the extent to which the most common form of network data (informant self-reports) can be considered to represent anything beyond the cognitions of the informants in question (see, for instance, Killworth and Bernard, 1976, 1979; Bernard and Killworth, 1977; Bernard et al., 1979, 1984; Freeman et al., 1987; Hildum, 1986; Romney and Faust, 1982; Romney et al., 1986). Indeed, Bernard et al. (1979) go so far as to posit that, "We are now convinced that cognitive data about communication can not be used as a proxy for behavioral data," a position which is rejected by researchers such as Romney et al. (1986) and Romney and Faust (1982). While the dust has not yet settled on this debate, the present state of knowledge within the field of network analysis strongly favors the position that informant self-reports contain considerable noise at the dyadic level (Bernard et al., 1984; Krackhardt and Kilduff, 1999). Whether the level of error observed is severe enough to warrant the assertion that self-report data is unusable for inference of interaction patterns appears to depend on a number of factors (including the time scale of the interaction, the application to which the data is to be put, and (arguably) the degree of optimism on the part of the researcher (Freeman et al., 1987; Freeman, 1992; Krackhardt and Kilduff, 1999; Bernard et al., 1979)), but unquestionably there is a non-trivial level of uncertainty inherent in self-report network data. The recognition that informants' reports of interaction are not in fact exact proxies for observed behavior has lead to two general approaches to the interpretation of such data. The first approach, which we shall here call the "classical" or "criterion/error" perspective, conceptualizes the data generation process in terms of a hypothesized "real" (or criterion) structure which is imperfectly reported by informants. This perspective is most clearly in line with the broader literature on informant accuracy (the notion of "accuracy" itself requiring the assumption of a criterion from which informant reports are assumed to deviate in some fashion), and is followed both by BKS and by later work seeking to identify the determinants of (in)accurate perception by socially embedded actors (see, for instance, Killworth and Bernard, 1976; Krackhardt, 1990; Calloway et al., 1993). The second approach, here referred to as the "cognitivist" perspective, has focused on identifying the determinants of actors' perceptions of social structure per se (e.g. Krackhardt, 1987a; Carley and Krackhardt, 1996; Freeman et al., 1987; Hammer, 1985; Hildum, 1986).

This perspective has tended to de-emphasize the role of the criterion structure, and indeed has in some cases questioned the relevance or even meaningfulness of such a construct (e.g. Krackhardt, 1987a). From a cognitivist point of view, the language of "error" and "accuracy" reflects a clear bias toward third-party observability (which may or may not be relevant to dyadic perceptions, particularly for private or tacit interactions) and non-attributive relations (Hildum, 1986). From the classical perspective, by contrast, these are critical considerations: inference regarding the criterion structures which are the objects of classical theorizing depends on the accuracy of informant reports, and the minimization of error is thus of central concern (Bernard et al., 1979).

It is beyond the scope of this paper to conduct a detailed consideration of the assumptions, applicability, and justification of the classical and cognitivist perspectives on self-report data. For our purposes, we shall adopt the language and assumptions of the classical approach, including the notions of the criterion graph, informant accuracy, and reporting error. Though the usefulness of that which follows for any given application is dependent upon the validity of these assumptions, we do not take them to be universally unproblematic; this issue is revisited in the discussion section below.

1.2. The enduring problem of error in social network analysis

While the informant accuracy question, then, has been considered from a variety of perspectives, at least four basic problems confront us with respect to dealing with error in social network analysis (from a classical point of view):

1. Determining the extent of error in existing data.
2. Determining the mechanisms by which error is produced.
3. Finding means of collecting higher quality data.
4. Minimizing and accounting for the uncertainty associated with existing data in network analyses.

All of these problems, of course, are of critical importance to social network analysis. Without knowing the degree of uncertainty with which we should regard present data, we have no way of evaluating the reliability (or even validity) of present or past work. Without knowing how or why errors are produced, we are at a loss to predict which data will be most heavily compromised. With no means of collecting clean data, we continue to be vulnerable to error and uncertainty, and without techniques for minimizing and assessing that uncertainty, we are unable to draw appropriate inferences. This paper, then, will touch at least briefly on all four issues, although the primary emphasis will be on the fourth problem.

1.3. A Bayesian approach

Given that social network analysis is heavily dependent upon informant reports, and that informants are known to err in their reporting of network ties, how are we to proceed? As noted above, a number of approaches are possible; in this paper, our primary interest is in the development of inferential techniques which can address the informant accuracy problem. Such techniques, clearly, must simultaneously accomplish two goals: first, they must allow

us to infer the criterion graph from informant reports; and second, they must allow us to infer the accuracy of each informant. Though solving both such problems at once would seem a difficult task, it is inescapable; inferences regarding informant accuracy will necessarily affect inferences regarding the criterion graph, and vice versa (see Batchelder and Romney (1988) and Romney et al. (1986) for additional treatment of this issue). This implies, further, that whatever approach is employed must scale effectively to high-dimensional inference problems. Assuming a criterion graph on N vertices, the number of informants (whose accuracy is in question) grows on order N and the number of ties in the criterion graph grows on order N^2; thus, the problem dimension is at least order N^2. Relative to this fairly large number of parameters, the data available is likely to be modest. For the best standard case in which all members of the network report on all ties—a cognitive social structure (CSS)—we have access to N observations per arc and N^2 observations per actor. While the total number of data points thus grows on order N^3, the small size of most CSSs provides us with relatively few observations per parameter. Our technique, then, must not only scale well, but must also be data-efficient. Methods which are justified only in the large-N limit are unlikely to be of practical value for this particular problem.

All of the above factors suggest the efficacy of a hierarchical Bayesian modeling approach to the network inference/informant accuracy problem (Gelman et al., 1995). Hierarchical Bayesian models can readily represent complex interrelated stochastic processes, scale well, and are not dependent on limit arguments for their justification. Furthermore, use of the Bayesian paradigm permits us to draw direct inferences regarding posterior probabilities, and grounds our inferential framework on an axiomatic basis.[2] Finally, the hierarchical Bayesian modeling framework readily facilitates expansion and modification of existing models to account for new information or to take advantage of the features of particular situations. Given these advantages, we here employ the aforesaid approach in examining the network inference/informant accuracy problem.

The models pursued here belong to a broader class of Bayesian models for binary data (Evans et al., 1995), although our development is informed by the particular concerns of network analysis. Likewise, it is important to note that their development builds on the growing tradition of statistical modeling within network analysis itself. In addition to the development of latent structural parameter models such as p^* (Wasserman and Pattison, 1996; Pattison and Wasserman, 1999; Robins et al., 1999), the past two decades have produced a growing battery of null hypothesis testing (Krackhardt, 1987b; Pattison et al., 2000; Anderson et al., 1999a), sampling (Frank, 1988; Marsden, 1990), and graph comparison (Krackhardt, 1988; Banks and Carley, 1994; Butts and Carley, 2001) techniques. Batchelder and Romney's 1988 consensus model is particularly similar in spirit to our approach (see also Romney et al., 1986); setting out to examine a closely related question of determining cultural "facts" from multiple informants' reports, they postulate an error model for individual accounts and derive point estimates for "facts" and individual competencies via maximum likelihood. The Romney/Batchelder notion of "competency" is a one-dimensional construct (borrowed from educational testing theory) which, in effect, treats errors symmetrically. (More precisely, Romney and Batchelder parameterize the problem in terms of "competency" and "bias" (in the direction of errors), but assume the latter to be absent

[2] For an examination of some of the foundational strengths of the Bayesian paradigm, see Robert (1994).

for their application.) By contrast, the parameterization employed here (see below) follows the logic of diagnostic testing in discriminating between false positive and false negative reports. While we regard this latter parameterization as more intuitive for this application, it should be noted that the two likelihoods can be made equivalent (up to a transformation) by freeing the Romney and Batchelder bias parameter. This implies, further, that the hierarchical prior structures explored here could be adapted to the Romney/Batchelder model, although this would potentially require modifying some of the latter's assumptions (e.g. the impossibility of perversely informative accounts). One way in which the present work differs from these approaches is that it emphasizes the evaluation of a variety of posterior quantities, in addition to point estimates. Thus, we are as concerned with quantifying (residual) uncertainty as with reducing it. Given that we analyze imperfect data in an imperfect world, we generally cannot claim to be completely certain of the quantities with which we work; a realistic approach to data analysis, then, cannot afford to simply brush problems of uncertainty under the proverbial rug. Arguably, however, much of traditional social network analysis has done exactly that: by failing to account for error in network data, we have put the quality of our inferences at risk. Even where the models presented here do not provide substantial uncertainty reduction (e.g. where only one observation is available per arc), they may still be of use in their ability to provide a concrete treatment of uncertainty in network data.

2. Bayesian models of the network inference/informant accuracy problem

As we have argued, an integrated, formal approach is required to simultaneously address the problems of network inference and informant accuracy. In this section, then, we develop a family of hierarchical Bayesian models (closely related to those of Evans et al. (1995)) which allow for inference on these two problems, and demonstrate the use of these models in determining posterior quantities of interest.

2.1. Notation and background assumptions

Before constructing a formal model of the data generation process, we first outline the basic notation which will be used in the discussion which follows. Given vertex set V and edge set E, the graph formed by E on V is denoted $G = (V, E)$; the functions $V(G)$ and $E(G)$ should be understood as returning the vertex and edge sets of G, respectively. Throughout this paper, we shall treat the criterion graph (G) as a digraph; although the term may be used generically to refer to any sort of graph (e.g. simple, signed, valued, etc.) which acts as the criterion ("true value") for informant reports, we will not treat these cases here.[3] Thus, the directed edges (or arcs) of the criterion are synonymous with ordered pairs (i, j) where $i, j \in V(G)$.

For the purposes of the present model, we assume that $V(G)$ is known, and that the vertices in question are uniquely labeled (i.e. are identifiable). Thus, we treat G as a random digraph of (known) order $N = |V(G)|$. For convenience, it will generally be more useful to

[3] Although we include the undirected case (with or without loops) as a special case involving missing data.

refer to G by way of its adjacency matrix, Θ, whose elements are defined as follows:

$$\Theta_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E(G) \\ 0 & \text{if } (i, j) \notin E(G) \end{cases} \tag{1}$$

Obviously, Θ is itself a random matrix whose elements serve as indicator variables for the arcs of G; it is the distribution of this matrix in which we are most interested. Our notation for observed graphs is similar. Let $(G_1, \ldots, G_M)$ be an ordered tuple of reported graphs. We then take Y to be the $N \times N \times M$ adjacency array with elements

$$Y_{ijk} = \begin{cases} 1 & \text{if } (i, j) \in E(G_k) \\ 0 & \text{if } (i, j) \notin E(G_k) \end{cases} \tag{2}$$

In the case of missing observations, we treat the corresponding cells as missing data. Where loops (i.e. (i, i) arcs) are undefined, the diagonals of Y and Θ are ignorably missing by design, and are simply omitted from all relevant calculations. Similarly (and without loss of generality), we represent undirected graphs by means of coding the upper or lower triangle of Θ as missing. For concreteness, we represent this missing data via the inclusion arrays I^{Θ} and I^{Y}. I^{Θ} is an $N \times N$ matrix whose elements are given by

$$I^{\Theta}_{ij} = \begin{cases} 1 & \text{if } \Theta_{ij} \text{ is observed} \\ 0 & \text{if } \Theta_{ij} \text{ is missing} \end{cases} \tag{3}$$

Note that inclusion/exclusion for I^{Θ} must be by design, since Θ is the parameter matrix for the criterion graph. (It is in this sense that we use the term "observed" above.) I^{Y} is similarly an $N \times N \times M$ array with elements

$$I^{Y}_{ijk} = \begin{cases} 1 & \text{if } Y_{ijk} \text{ is observed and } I^{\Theta}_{ij} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Exclusion in I^{Y} obviously need not be by design: arcs which are missing for more conventional reasons may be excluded as well. Throughout this paper, it will be assumed that inference for Θ and other relevant parameters is ignorable with respect to the inclusion matrices. That is, we assume that inference for parameters of interest depends only on the observed-data likelihood, and not on the inclusion structure itself.[4] Some discussion of sampling designs which satisfy this criterion is provided below; note, however, that a complete graph census omitting only undefined arcs (e.g. loops) immediately qualifies as an ignorable design.

Prior to discussing the data generation model, a final note is in order regarding notation for random variables and their distributions. For a random variable x, the function $p(x)$ should be interpreted either as the probability density function or as the probability mass function of x, for continuous or discrete x (respectively). The terms "density" and "distribution" are used somewhat loosely to refer to both of the above, where there is no danger of confusion. Conditional notation (e.g. $p(x|y)$ for the density of x given y) is used to distinguish prior and

[4] Note that we may use the inclusion structure as a convenience in calculating the observed-data likelihood; the requirement is rather that it does not play a role in our inference per se.

posterior densities, and likelihoods are similarly given in terms of the probability density of observations conditional on model parameters. When making reference to standard parametric densities, an abbreviation of the name will be used. Thus $p(x|\alpha, \beta) = \text{Unif}(x|\alpha, \beta)$ implies that x is uniformly distributed conditional on parameters α and β, a situation which may be equivalently denoted $x \sim \text{Unif}(\alpha, \beta)$ (where $\sim$ is read "is distributed as"). These abbreviations will be defined where initially deployed within the text.

2.1.1. General informant accuracy model

As argued above, one cannot address the problem of informant accuracy without having some prior conception of the manner in which the hypothesized criterion structure is related to the pattern of informant accounts. This is not a trivial matter: the relationship in question has been argued to be mediated by cognitive mechanisms, instrument design, informant experience, and even the informant's position in the social structure itself (Freeman, 1992; Freeman et al., 1987; Carley and Krackhardt, 1996; Krackhardt, 1990; Hildum, 1986; Krackhardt and Kilduff, 1999). Plainly, one can propose models with varying degrees of sophistication, which take into account a greater or lesser number of influences on accuracy, and which model those influences in more or less sophisticated ways. In this paper, our primary purpose is to introduce a reasonably simple family of models which are easily utilized in network research; our secondary purpose is to formulate these models in such a way as to make them as compatible as possible with the prior empirical work on informant accuracy within the context of social network analysis. For this reason, we shall utilize a general informant accuracy framework which follows that implicitly utilized (we shall argue) by Bernard, Killworth, Sailer, and others in their informant accuracy studies. Although this framework is reasonably flexible, it is not asserted that this approach will be optimal in all cases; as will be discussed, some applications may be better suited by alternative formulations. If we are to treat informant accuracy in such a way as to maintain compatibility with prior work, it behooves us to consider the models which have been (explicitly or implicitly) invoked in the past. Fortunately, the most common approach has been described fairly clearly by BKS. With respect to the basic structure of the problem, the version articulated here is quite typical: "Social structure is assumed to be built up out of the interaction of the members of the group. Then, the plausible leap is made whereby the answers to [a] sociometric question reflect the pattern of interactions" (Bernard et al., 1979). Thus, it is assumed that there exists a criterion network which is formed from "actual" interactions between actors, and accuracy is measured with respect to the degree that informant responses on sociometric instruments match this network. More explicitly, BKS state, "At its simplest level, network data are 'accurate' if, when *i* says he talked to *j* by some amount, then he did" (Bernard et al., 1979). Though this statement may seem to be self-evident at first blush, its implications are non-trivial. For instance, the BKS model requires that the "actual" network be in some sense verifiable—and hence defined—beyond the individual perceptions or testimonials of those involved in the proposed relation.[5] This sets the scope

[5] BKS, indeed, repeatedly question the meaningfulness of any structure which does not have this feature, even going so far as to argue that networks are "real" insofar as they "have some external correlate like performance in problem solving" (Killworth and Bernard, 1979). It is not our intention to make arguments for or against this position in this paper, but the issue is obviously one of great relevance to the continued theoretical and methodological development of network analysis.

of the accuracy question (from this perspective) in such a way as to exclude purely ascriptive relations (e.g.: individual perceptions of influence, love, or power; reputation and prestige;[6] and some notions of friendship[7]), at least on a first-order basis.[8] As the BKS statement makes clear, the focus here is not on accounting for informant responses to sociometric instruments, but rather on determining the degree to which these responses are consistent with a pre-defined observable (but unobserved) relation. Although the former has been considered to be an interesting question in its own right at least since Krackhardt's introduction of the cognitive social structure concept (Krackhardt, 1987a), this is not our primary interest in this case.

The base assumption which is made, then, is that each observer is exposed to the underlying "real" network and reports observed ties which are, with some probability, erroneous. It should be emphasized that we are unconcerned at present with whether the errors in question are the result of inaccurate observation by the informant, errors in the cognitive coding of the relation, difficulties with retrieval of this information upon exposure to the appropriate instrument, or informant malfeasance. These error sources are folded together in the present model, and we are concerned only with the aggregate result. As there are by assumption two states which may be taken by each arc (present and absent), there are clearly two ways in which observers can report incorrectly: an observer may report present ties as being absent (false negatives) or may report absent ties as being present (false positives). We then model the data generation process via the following Bernoulli mixture,

$$p(Y_{ijk}|\Theta_{ij}, e^+, e^-) = \begin{cases} B(Y_{ijk}|1-e^-) & \text{if } \Theta_{ij} = 1 \\ B(Y_{ijk}|e^+) & \text{if } \Theta_{ij} = 0 \end{cases} \tag{5}$$

where Y and Θ are defined as per the previous section, where e^+ and e^- represent the probabilities of false positives and false negatives (respectively) for the ijkth arc report, and where B is the Bernoulli density. The above provides the core likelihood for a single observer/arc/criterion network combination—it provides the probability density associated with a given report, conditional on the (uncertain) parameters. In general, of course, we are interested in drawing inferences regarding these parameters, and we will need to draw upon multiple informant reports regarding multiple arcs. In the sections which follow, we will build on Eq. (5) to form models which can account for these complications.

2.1.2. Assignment of network priors

To infer the criterion, we must know two things: the likelihood for informant reports, and the (joint) prior distribution of the model parameters. While any number of assumptions are possible regarding this latter distribution, it seems reasonable—as a starting point, at least—to treat the criterion graph and error parameters as a priori independent. Likewise, we will here make the simplifying assumption that our prior knowledge of the criterion

[6] Note that unlike dominance and power, which can be fairly readily defined beyond their ascriptions (though *perceptions* of such are always ascriptive), reputation and prestige are by nature ascribed relations.

[7] See, for instance, Carley and Krackhardt (1996).

[8] E.g. one can still ask whether a given actor's account of the ascriptions of others matches their actual ascriptions, but this is a second-order rather than a first-order question.

graph can be roughly summarized by the distribution

$$p(\Theta) = \prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij}) \tag{6}$$

$$p(\Theta) = \prod_{i=1}^{N}\prod_{j=1}^{N} ((1 - I_{ij}^{\Theta}) + I_{ij}^{\Theta} B(\Theta_{ij}|\Phi_{ij})) \tag{7}$$

i.e. the Bernoulli graph distribution with hyperparameter matrix Φ (Wasserman and Faust, 1994) under inclusion matrix I^{Θ}. The Bernoulli graph distribution is a natural baseline model for network structure, and (particularly when Φ is chosen based on external covariates) can incorporate considerable prior information. On the other hand, as this distribution does not incorporate dependency constraints between arcs (conditional on Φ), it may also be employed as an "overdispersed" prior for purposes of robustness. While there may be special circumstances in which the Bernoulli prior is clearly inappropriate, it is expected that these will be uncommon; posterior inference with arbitrary exponential random graph priors will be briefly outlined below. Assuming, then, a Bernoulli graph prior, we must specify values for Φ which reflect our initial information regarding the criterion graph. This must, of necessity, depend on the particular problem at hand, the existing literature regarding said problem, etc. Nevertheless, it is useful to provide some basic strategies for assigning network priors, which may be utilized in a range of contexts. Two heuristics, then, are here presented which are expected to have wide applicability; each researcher, however, should be careful to select network priors which are accurate depictions of his or her prior information, and should avoid blind reliance on pre-packaged choices.

The first, and perhaps most obvious, heuristic for assignment of network priors is that of an uninformative distribution on the arc set of G. Such a prior distribution is given simply by

$$\Phi_{ij} = 0.5 \forall i, j \in (1, \ldots, N) \tag{8}$$

and corresponds to a uniform distribution on the set of all digraphs of order N. This prior is obviously somewhat attractive in that it has a clear interpretation, and that it does not depend on any idiosyncratic knowledge of the researcher. In general, however, an uninformative prior on the set of digraphs is a poor reflection of one's prior information, and better options are available; researchers should question whether or not they are in fact completely ignorant as to the structure of the criterion graph before selecting such a prior.

One specific example of the weakness of an uninformative network prior is the matter of network density. While the uniform distribution over digraphs favors graphs of density approximately equal to 0.5,[9] most social structures (particularly large ones) are considerably less dense. If a researcher has reason to believe that the relation he or she is studying is of a particular density, it behooves him or her to take this into account when selecting a prior for the criterion graph.[10] One possibility, then, for a somewhat more informative network prior is to chose a distribution of the form

$$\Phi_{ij} = d \forall i, j \in (1, \ldots, N) \tag{9}$$

[9] For reasons of combinatorics: there are far more structures of moderate density than of extreme density.

[10] For a discussion of the effects of size and density on network inference, see Anderson et al. (1999a).

where d is the median density of a set of other, similar networks examined in past research. While this is obviously only one of many possible density-based priors, it is clearly more informative than the assumption of a uniform distribution, and is nevertheless diffuse enough to avoid extreme sensitivity to initial assumptions. In situations for which more information is available, researchers should not hesitate to use it: priors including homophily or distance effects, for instance, might be reasonable in many situations, as might the edgewise marginal posterior distribution of the same network at an earlier point in time (see below). The topic of network prior selection is complex enough to warrant a detailed treatment on its own, and it is hoped that future work in this area will develop more sophisticated strategies for exploiting prior knowledge that can be discussed within the bounds of this paper.

2.2. A simple model for fixed, known error probabilities

Before we begin a consideration of the more complex cases of network inference in the presence of unknown informant inaccuracies, it behooves us to begin with a simpler model. As an introduction to our basic modeling framework, then, we shall first develop a network inference model in which we assume that our error probabilities are both fixed and known. Clearly, this does not reflect the context of most network research: we do not, in general, know the probability of error within our data to within an arbitrary degree of certainty. Nevertheless, there may be special cases—such as, for instance, automated data collection procedures whose error rates may be determined exogenously—for which the simple model may be applicable. Our reason for introducing it, however, is as much to elucidate the concepts involved as to provide a practical tool for social research.

2.2.1. Assignment of priors

Within the simple network inference model, it is assumed that only the criterion graph is uncertain; therefore, priors are assigned only to the network parameter matrix, Θ. As indicated, the prior distribution of the criterion graph is determined via the hyperparameter matrix Φ. The specific form, then, is as per Eq. (7). Choice of the values of Φ is discussed in the above section on network priors; as mentioned, these values should reflect the researcher's prior knowledge regarding the structure of the criterion graph.

2.2.2. Assumed likelihood

As we are assuming in this case that our error probabilities are fixed and known, the base likelihood for an arc observation follows straightforwardly from Eq. (5) and the Y inclusion structure:

$$p(Y_{ijk}|\Theta_{ij}, e^+, e^-) = (1 - I^Y_{ijk}) + I^Y_{ijk}((1 - \Theta_{ij})(Y_{ijk}e^+ + (1 - Y_{ijk})(1 - e^+)) + \Theta_{ij}(Y_{ijk}(1 - e^-) + (1 - Y_{ijk})e^-)) \quad (10)$$

Observe that this is just the mixture from Eq. (5) (written algebraically), "filtered" by the inclusion structure. (Our notation suppresses the dependence on I^Y, which is assumed to be implicit; we will suppress dependence on any relevant hyperparameters for the same reason.) In particular, one can readily verify that the likelihood of any missing observation is 1, that the likelihood of a true positive is $1 - e^-$, etc. To find the joint likelihood of the data, we

invoke the previously stated assumption of conditional independence of arc observations; this implies that the joint likelihood is the product of the individual arc likelihoods. Formally, the joint likelihood of the simple model is then given by

$$p(Y|\Theta, e^+, e^-) = \prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e^+, e^-) \tag{11}$$

where Y is the observed data matrix and $p(Y_{ijk}|\Theta_{ij}, e^+, e^-)$ is the arc likelihood of Eq. (10). (Again, we have suppressed dependence on the inclusion structure.)

2.2.3. Computation of the posterior

Given the above, we can identify the posterior to within a normalizing constant by using the standard result that $p(\theta|y) \propto p(\theta)p(y|\theta)$ for random variables y and θ (i.e. the proportional form of Bayes' theorem). Thus, the unnormalized posterior of Θ given Y and the (known) error rates is given by

$$p(\Theta|Y, e^+, e^-) \propto p(Y|\Theta, e^+, e^-)p(\Theta) \tag{12}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e^+, e^-)\right)\left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\right) \tag{13}$$

$$= \prod_{i=1}^{N}\prod_{j=1}^{N}\left(p(\Theta_{ij})\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e^+, e^-)\right) \tag{14}$$

$$= \prod_{i=1}^{N}\prod_{j=1}^{N} q(\Theta_{ij}|Y_{ij\cdot}, e^+, e^-) \tag{15}$$

where $q(\Theta_{ij}|Y_{ij\cdot}, e^+, e^-)$ is the unnormalized posterior marginal of Θ_{ij}. Thus, we see that the joint posterior in this case is factorable by arc, a fact which we can use to our advantage. By applying Bayes' theorem to each arc independently, we may derive an explicit expression for the exact posterior. This is, in essence, possible because we do not allow inference on any given arc to affect our inference regarding any other arc; no information is shared across observations, and hence we can subdivide the problem into a number of simple Bernoulli mixtures. The posterior probability for a given arc, then, is

$$p(\Theta_{ij}|Y_{ij\cdot}, e^+, e^-) = (1 - I_{ij}^{\Theta}) + I_{ij}^{\Theta}\frac{p(Y_{ij\cdot}|\Theta_{ij}, e^+, e^-)p(\Theta_{ij})}{\sum_{\ell=0}^{1} p(Y_{ij\cdot}|\Theta_{ij} = \ell, e^+, e^-)p(\Theta_{ij} = \ell)} \tag{16}$$

$$= (1 - I_{ij}^{\Theta}) + I_{ij}^{\Theta} B\left(\Theta_{ij}\left|\frac{\Phi_{ij}\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|1 - e^-)]}{\begin{array}{l}\Phi_{ij}\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|1 - e^-)] \\ \quad + (1 - \Phi_{ij})\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|e^+)]\end{array}}\right.\right) \tag{17}$$

and the (normalized) joint posterior of the criterion graph is hence

$$p(\Theta|Y, e^+, e^-) = \prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij}|Y_{ij\cdot}, e^+, e^-) \tag{18}$$

The above, of course, provides the criterion posterior for M observers. Should additional observations become available, inference may still be drawn in the usual fashion (i.e. by repeating the above computation with the initial posterior as prior and with the new observations as Y). Furthermore, the assumption that e^+ and e^- be identical for all observations may be trivially relaxed in the simple case. To accomplish this, we replace the single values e^+, e^-, above, with e^+_{ijk}, e^-_{ijk} for each arc observation. So long as these edgewise error parameters are known, there is no requirement that they be identical for all arcs; unfortunately, this result is limited to the simple model (as we shall see).

In considering the posterior probabilities of Eqs. (17) and (18), it is helpful to make a few observations regarding the effect of prior knowledge. (For simplicity, we will assume here that no missing data is present, and we will limit ourselves to the single observer case in which Y is an $N \times N$ matrix.) As our error rates e^+ and e^- simultaneously approach zero, then clearly $\Theta_{ij} \to Y_{ij}$; by turns, for $e^+, e^- \to 0.5$, $p(\Theta|Y) \to p(\Theta)$. This expresses the fact that the amount of information carried by Y depends on the error rates: at the extremes, no error implies that our observations must logically dominate our prior information, and symmetric error rates of 50% imply that observations carry no information whatsoever. More broadly, the information carried by a given edge observation is related to $1 - (e^+ + e^-)$, with increasingly positive values indicating increasing information in the direction of the observation, and increasingly negative values indicating increasing information in the opposite direction (i.e. perverse or "negatively informative" observations). This general relationship is true regardless of whether $e^- = e^+$, although unequal error rates imply that the magnitude of effects from tie present and tie absent observations will differ accordingly.

On a practical note, it should be mentioned that the above analytical results imply that the joint posterior of the criterion graph need not be used directly for most applications (though its computation is obviously quite straightforward where required). To draw from the posterior of the criterion graph in this case, one need only draw each Θ_{ij} independently with probabilities as given by Eq. (17) above. This makes estimation of posterior graph properties (see Section 2.6 below) particularly easy for the simple error model; this ease, however, comes at the expense of any inference regarding error probabilities (which are presumed known on an a priori basis) and of any shared information across arcs. The simple error model, then, is generally useful only when the accuracy of observers can he established ex ante, a condition which is rarely encountered in a typical social network analysis setting. It nevertheless serves as a reasonable prelude to the more sophisticated models which follow.

2.3. Pooled error probabilities (single observer model)

Having considered a simple model of network inference in the special case for which error probabilities are fixed and known, we can now proceed to a more sophisticated—and useful—model. In particular, one reasonable elaboration of the fixed error probability model is the case in which error probabilities are uncertain, but assumed to be constant

across observations.[11] Such a pooled error probability model readily represents a situation in which a single observer or informant (e.g. an ethnographer) provides reports on all ties within a given network (e.g. the observers in the data sets of Bernard et al. (1979)); we may be willing to presume that his or her accuracy is more or less constant across cases, but we are uncertain regarding the degree of accuracy itself.

2.3.1. Assignment of priors

Unlike the simple error model, the single observer model treats error probabilities as uncertain; hence, we must assign priors to them. While a wide range of distributions are possible here, we have elected to represent the error probabilities e^+ and e^- as being drawn independently from two Beta distributions. Specifically:

$$e^+ \sim \text{Beta}(\alpha^+, \beta^+) \tag{19}$$

$$e^- \sim \text{Beta}(\alpha^-, \beta^-) \tag{20}$$

where $\text{Beta}(x|\alpha, \beta) = ((\Gamma(\alpha+\beta))/(\Gamma(\alpha)\Gamma(\beta)))x^{\alpha-1}(1-x)^{\beta-1}$ for $x \in [0, 1]$. The choice of the Beta for the form of the error prior requires some justification; indeed, it may not be appropriate in all cases. The Beta distribution is the conjugate prior for the Binomial likelihood, with the uninformative prior distribution on [0, 1] being a special case.[12] As a result, the outcome of any set of success/failure Bernoulli experiments which begins with an uninformative prior distribution will be Beta distributed, and the Beta is thus a logical choice when one's prior information can be summarized in this form. In general, prior knowledge regarding informant accuracy is likely to come from error rates derived from earlier studies (e.g. BKS), and hence it is not unreasonable to suppose that, for many researchers, the Beta will be an appropriate representation by this argument. Those drawing prior knowledge from other sources may need to consider alternative forms; these are not pursued within this paper. Given the choice of Beta distributions for e^+ and e^-, there remains the question of selecting the four hyperparameters α^+, β^+, α^-, and β^-. While these could, in principle, be themselves drawn from a hyperprior distribution, we here treat them as known from prior data. As noted, the choice $\alpha = \beta = 1$ provides an uninformative prior for the Beta distribution; however, it is *strongly* recommended that researchers avoid this choice of prior. The reason for this is simple: the assumption of a uniform distribution of error parameters is highly unreasonable for most applications (given previous research in this area) and leads to highly counterintuitive (and improbable) inferences. Note, for instance, that (by Bayes theorem) the condition $e^+ + e^- > 1$ leads to a condition of *perverse inferences*, in which informant testimony causes one to update one's belief in the *opposite* direction of the report. Clearly, this is an unlikely event: even at their worst, it is hard to imagine that most informants' reports would be *negatively* informative![13] (This is not strictly impossible, however, particularly

[11] Recall that constant overall error rates were initially assumed for the fixed error probability model, but that we were able to relax this requirement due to our ability to decompose the problem into a series of arc inferences.

[12] In fact, there are no fewer than three uninformative priors for this particular problem: the uniform distribution ($\alpha = \beta = 1$), the Jeffrey's prior ($\alpha = \beta = 1/2$), and the improper prior which is uniform in the natural parameter of the exponential family ($\alpha = \beta = 0$) (Gelman et al., 1995). We assume the uniform distribution as an uninformative prior unless otherwise indicated.

[13] This has other implications as well: for instance, the posterior construction of two basic scenarios such that in one scenario informant reports are nearly all versed in implication.

where malfeasance cannot be ruled out.) Under a uniform prior, the a priori probability of such an occurrence is 0.5—unacceptably high for most applications. In general, then, α and β parameters should be chosen so as to cause the distribution of error probabilities to remain sensible, and to prevent the perverse inference condition from being highly probable. In examining a number of stylized facts from the BKS studies (i.e. apparent tendencies in error rates across relations), a prior resembling Beta(3, 9) for both false negative and false positive error parameters has thusfar seemed reasonable for communication-like relations; individual researchers, however, should base their choices of hyperparameters on the particular data available. Given the choice of priors for the two error parameters, the form of the network prior is as for the simple error model, specifically Eq. (7). (Note that we presume that the joint prior for Θ, e^+, and e^- is factorable, i.e. that $p(\Theta, e^+, e^-) = p(\Theta)p(e^+)p(e^-)$.) As the issues in selecting the network hyperparameters are not distinct in this case, the reader is directed to the previous discussion regarding choice of Φ.

2.3.2. Assumed likelihood

While our model has been changed by the introduction of uncertainty regarding our error parameters, the likelihood of the data is unaltered from our previous formulation. Thus, our arc likelihood is given by Eq. (10) and the joint likelihood for the entire data set is, as before, the product of the arc likelihoods (Eq. (11)).

Note that while the assumption of uncertainty in error parameters does not alter our likelihood, it will nevertheless affect posterior inference (as we shall see presently). Intuitively, the reason for the former is that the likelihood of the observed data already conditions on the error parameters, and hence treats them as "fixed". The role of stochasticity in errors, then, is seen in the computation of the posterior.

2.3.3. Computation of the posterior

In the simple error model, we were able to express the posterior of the criterion graph in very simple terms; in the single observer model, this is complicated by the stochasticity of the error parameters, e^+ and e^-. To derive the unnormalized joint posterior we invoke the proportional form of Bayes theorem, as follows:

$$p(\Theta, e^+, e^-, Y) \propto p(Y|\Theta, e^+, e^-)p(\Theta)p(e^+)p(e^-) \tag{21}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e^+, e^-)\right) \times \left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\right) \text{Beta}(e^+|\alpha^+, \beta^+)\text{Beta}(e^-|\alpha^-, \beta^-) \tag{22}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\left(p(\Theta_{ij})\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e^+, e^-)\right)\right) \text{Beta}(e^+|\alpha^+, \beta^+)\text{Beta}(e^-|\alpha^-, \beta^-) \tag{23}$$

Note that (unlike in Eq. (15)) we cannot factor the posterior further, due to the interaction between the individual edge likelihoods and the error parameters. Now the impact of uncertainty in the error parameters becomes clear: we cannot consider the probability of the criterion graph without considering the probability of the error parameters which give rise to that inference, and neither can we consider the probability of a particular pair of error parameters without taking into account the probability of the structural inference such a choice would induce. As asserted, the two problems are inseparable; this intuition is laid bare in the form of the joint posterior.

Given that Eq. (21) does not lend itself to direct computation, how may we employ it in practice? One approach, which is facilitated by the form of the posterior, is to simulate posterior draws from the joint posterior using a Gibbs sampler, and to in turn use these draws to estimate posterior quantities of interest. The Gibbs sampler—a Markov Chain Monte Carlo (MCMC) method—works by taking a series of draws from the full conditionals of the posterior, constructing a Markov chain whose equilibrium distribution converges to that of the joint posterior. Details concerning the use of MCMC techniques in a Bayesian context can be found in Gamerman (1997); our discussion here will focus exclusively on the conditional distributions which are necessary for its implementation.

The first of the conditional distributions we require is that of the criterion graph, conditional on the realizations of the two error parameters and of the data matrix. As the partial factorization of Eq. (23) indicates, this conditional density reduces to the simple product of the arc posteriors; these were derived in Eq. (17) for the simple error model. The joint conditional of the criterion graph, then, is given by

$$p(\Theta|e^+, e^-, Y) = \prod_{i=1}^{N}\prod_{j=1}^{N} \left[(1 - I^{\Theta}_{ij}) + I^{\Theta}_{ij} \times B\left(\Theta_{ij} \left| \frac{\Phi_{ij}\prod_{k=1}^{M}[(1-I^{Y}_{ijk})+I^{Y}_{ijk}B(Y_{ijk}|1-e^-)]}{\Phi_{ij}\prod_{k=1}^{M}[(1-I^{Y}_{ijk})+I^{Y}_{ijk}B(Y_{ijk}|1-e^-)] + (1-\Phi_{ij})\prod_{k=1}^{M}[(1-I^{Y}_{ijk})+I^{Y}_{ijk}B(Y_{ijk}|e^+)]} \right) \right] \tag{24}$$

Note that this expression is arc factorable, and hence we can readily simulate Θ by drawing from

$$\Theta_{ij}|e^+, e^-, Y \sim B\left(\Theta_{ij} \left| \frac{\Phi_{ij}\prod_{k=1}^{M}[(1 - I^{Y}_{ijk}) + I^{Y}_{ijk}B(Y_{ijk}|1 - e^-)]}{\Phi_{ij}\prod_{k=1}^{M}[(1 - I^{Y}_{ijk}) + I^{Y}_{ijk}B(Y_{ijk}|1 - e^-)] + (1 - \Phi_{ij})\prod_{k=1}^{M}[(1 - I^{Y}_{ijk}) + I^{Y}_{ijk}B(Y_{ijk}|e^+)]} \right) \right. \tag{25}$$

independently for all arcs such that $I^{\Theta}_{ij} = 1$. (All other arcs should properly be coded as missing, and obviously do not enter into this calculation.)

For the conditional probability of e^+, we begin by noting an alternative factorization of the joint posterior density:

$$p(\Theta, e^+, e^-|Y) \propto \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e^+, e^-)\right)\left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\right) \times \text{Beta}(e^+|\alpha^+, \beta^+)\text{Beta}(e^-|\alpha^-, \beta^-) \tag{26}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M}[(1 - I^Y_{ijk}) + I^Y_{ijk}(\Theta_{ij}B(Y_{ijk}|1 - e^-) + (1 - \Theta_{ij})B(Y_{ijk}|e^+))]\right) \times \left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\right)\text{Beta}(e^+|\alpha^+, \beta^+)\text{Beta}(e^-|\alpha^-, \beta^-) \tag{27}$$

which, by the Boolean properties of Θ and I^Y, can be rewritten as

$$= \left(\left[\prod_{(i,j,k):\Theta_{ij}=1, I^Y_{ijk}=1} B(Y_{ijk}|1 - e^-)\right]\text{Beta}(e^-|\alpha^-, \beta^-)\right) \times \left(\left[\prod_{(i,j,k):\Theta_{ij}=0, I^Y_{ijk}=1} B(Y_{ijk}|e^+)\right]\text{Beta}(e^+|\alpha^+, \beta^+)\right)\left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\right) \tag{28}$$

Observe that this expression allows us to separate observations into true positive/false negative and true negative/false positive categories. When conditioning on Θ, then, we are left with a factorable joint conditional:

$$p(e^+, e^-|\Theta, Y) \propto \left(\left[\prod_{(i,j,k):\Theta_{ij}=1, I^Y_{ijk}=1} B(Y_{ijk}|1 - e^-)\right]\text{Beta}(e^-|\alpha^-, \beta^-)\right) \times \left(\left[\prod_{(i,j,k):\Theta_{ij}=0, I^Y_{ijk}=1} B(Y_{ijk}|e^+)\right]\text{Beta}(e^+|\alpha^+, \beta^+)\right) \tag{29}$$

which establishes the conditional independence of e^+ and e^- given Θ. Thus, to obtain the conditionals for the two error parameters, we need only focus on their respective terms in Eq. (29).

Beginning with the false positive probability, we note that

$$p(e^+, e^-|\Theta, Y) \propto \left[\prod_{(i,j,k):\Theta_{ij}=0, I^Y_{ijk}=1} B(Y_{ijk}|e^+)\right]\text{Beta}(e^+|\alpha^+, \beta^+) \tag{30}$$

$$\propto \left[(e^+)^{\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}(1-\Theta_{ij})Y_{ijk}}(1-e^+)^{\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}(1-\Theta_{ij})(1-Y_{ijk})}\right] \times (e^+)^{\alpha^+-1}(1-e^+)^{\beta^+-1} \tag{31}$$

$$= (e^+)^{\alpha^+ + \left(\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}(1-\Theta_{ij})Y_{ijk}\right)-1} \times (1-e^+)^{\beta^+ + \left(\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}(1-\Theta_{ij})(1-Y_{ijk})\right)-1} \tag{32}$$

Eq. (32) is recognizable as the unnormalized form of a Beta distribution, and thus we may take conditional posterior draws of e^+ via

$$e^+|e^-, \Theta, Y \sim \text{Beta}\left(\alpha^+ + \sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}(1-\Theta_{ij})Y_{ijk}, \beta^+ + \sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}(1-\Theta_{ij})(1-Y_{ijk})\right) \tag{33}$$

Note that the parameters of the conditional posterior are simple counts of successful and unsuccessful identifications of the non-existence of ties, combined with the prior parameters. As it happens, this is precisely what we would expect: conditional on $\Theta_{ij} = 0$, edge reports are binomial, and the Beta density forms a conjugate prior for the binomial likelihood. The same logic, then, would be expected to hold for the false negative probability:

$$p(e^-|e^+, \Theta, Y) \propto \left[\prod_{(i,j,k):\Theta_{ij}=1, I^Y_{ijk}=1} B(Y_{ijk}|1-e^-)\right] \text{Beta}(e^-|\alpha^-, \beta^-) \tag{34}$$

$$\propto \left[(e^-)^{\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}\Theta_{ij}(1-Y_{ijk})}(1-e^-)^{\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}\Theta_{ij}Y_{ijk}}\right] \times (e^-)^{\alpha^--1}(1-e^-)^{\beta^--1} \tag{35}$$

$$= (e^-)^{\alpha^- + \left(\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}\Theta_{ij}(1-Y_{ijk})\right)-1}(1-e^-)^{\beta^- + \left(\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}\Theta_{ij}Y_{ijk}\right)-1} \tag{36}$$

and we find that a similar pattern indeed holds. Here, our binomial is determined by the arc reports associated with arcs for which $\Theta_{ij} = 1$, and the corresponding Beta distribution can be written

$$e^-|e^+, \Theta, Y \sim \text{Beta}\left(\alpha^- + \sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}\Theta_{ij}(1-Y_{ijk}), \beta^- + \sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^M I^Y_{ijk}\Theta_{ij}Y_{ijk}\right) \tag{37}$$

In both cases, one would simulate draws from the respective posterior conditional distributions by taking independent Beta draws with the parameters given by Eqs. (33) and (37).

This is readily accomplished using most statistical computing packages, either directly or by drawing from the two distributions using the following two step procedure:

1. **procedure** Draw from Beta(a, β)
2. Draw $a \sim \chi^2_{2\alpha}$, $b \sim \chi^2_{2\beta}$
3. **return** $a/(a+b)$

(χ^2 deviates themselves can be drawn in a variety of ways, including as sums of squared standard uniform deviates. See Evans et al. (2000) or Gentle (1998) for details.)

We now have all of the necessary ingredients for posterior simulation via the Gibbs sampler. By iteratively drawing from the conditional posteriors of the criterion graph (Θ), e^+, and e^-, one can obtain a series of deviates which, in the limit, converge to the true joint posterior. These draws may, in turn, be used to simulate posterior quantities and/or to draw from posterior predictive distributions. While a thorough discussion of this procedure is beyond the scope of this paper, the following provides a simple description of the general algorithm:

1. **procedure** Draw from Θ, e^+, $e^-|Y$ (Single Observer Model)
2. Draw $\Theta^{(1)}$ from $p(\Theta)$
3. Draw $e^{+(1)}$ from $p(e^+)$
4. Draw $e^{-(1)}$ from $p(e^-)$
5. $i := 2$
6. **repeat**
7. Draw $\Theta^{(i)}$ from $p(\Theta|e^{+(i-1)}, e^{-(i-1)}, Y)$
8. Draw $e^{+(i)}$ from $p(e^+|\Theta^{(i)}, e^{-(i-1)}, Y)$
9. Draw $e^{-(i)}$ from $p(e^-|\Theta^{(i)}, e^{+(i)}, Y)$
10. $i := i + 1$
11. **until** $\Theta^{(\cdot)}, e^{+(\cdot)}, e^{-(\cdot)} \sim \Theta, e^+, e^-|Y$
12. **return** $\Theta^{(\cdot)}, e^{+(\cdot)}, e^{-(\cdot)}$

It should be cautioned that, depending on the nature of the stochastic process, many iterations may be required prior to convergence (the point at which the sample draws adequately approximate the true distribution).[14] The length of this initial period (known as the "burn-in" time of the chain) may be assessed using a variety of diagnostics; see Gamerman (1997) for an accessible review. Draws taken prior to burn-in are generally dependent upon the starting position of the chain, and hence are typically discarded. Numerous other considerations (including choice of starting value, multiple versus single chain methods, etc.) may also enter into implementation of a MCMC routine, and the reader is referred to the above references for details.

2.4. Multiple error probabilities (multiple observer model)

Having developed models for both fixed and pooled error probabilities, we are now ready to proceed to the more general case in which our network data is generated by a process

[14] Technically, convergence occurs only asymptotically, but a finite-sample approximation is satisfactory for most applications.

involving multiple uncertain error probabilities. The canonical example of such a situation is that of a cognitive social structure, in which each actor within the structure reports the full set of relations among actors (presumably with uncertain error probabilities which vary by informant). Such a data set provides us with multiple observations across both arcs and actors; thus, we have a fair amount of leverage in drawing inferences regarding both informant accuracy and the criterion graph. Alternately, such data may emerge from a random sample of informants recruited by the researcher, or from the combined accounts of multiple third-party observers. By allowing each observer's pattern of error generation to vary, we greatly expand our ability to integrate network data from a variety of sources.

2.4.1. Assignment of priors

The priors required for the multiple observer model are, as one would expect, a simple generalization of those employed in the single observer model. Whereas before we had only two sets of hyperparameters, here we have twice as many sets as the number of observers; the e^+ priors, for instance, under the assumptions of conjugacy and conditional independence, are given by

$$e_i^+ \sim \text{Beta}(\alpha_i^+, \beta_i^+), \quad i \in \{1, 2, \ldots, M\} \tag{38}$$

$$p(e^+) = \prod_{i=1}^{M} p(e_i^+) \tag{39}$$

Note that the hyperparameters are not required to be identical for all actors. Indeed, if individuating information (e.g. from past observation of a particular informant) is available, it should clearly be employed. Again with the same assumptions, we take the form of the e^- priors to be identical:

$$e_i^- \sim \text{Beta}(\alpha_i^-, \beta_i^-), \quad i \in \{1, 2, \ldots, M\} \tag{40}$$

$$p(e^-) = \prod_{i=1}^{M} p(e_i^-) \tag{41}$$

It is worth pointing out, with respect to the above, that implicit in our assumptions is the requirement that inferences regarding individual error parameters affect each other only indirectly via the estimation of the criterion graph. A reasonable extension of this model might relax this assumption by taking the hyperparameters for the individual actors as being drawn from a hyperprior distribution; one candidate for such a distribution would be a Gamma, which has a number of desirable properties in such a role. An extension of this type, of course, would add yet another layer to the hierarchical model, and will not be pursued here. Nevertheless, it is a promising direction for future research.

In addition to the choice of prior distributions for the error parameters, we are still left with the usual problem of defining priors for the criterion graph. The form for the criterion prior is unchanged from its previous incarnations, and follows Eq. (7).

2.4.2. Assumed likelihood

The assumed likelihood for the multiple observer medel is a straightforward extension of the likelihood for the single observer model. Each observation is, as usual, a Bernoulli mixture, with the distinction in this case being the use of separate error probabilities for each observer. Following Eq. (10), then, the arc likelihood is given by

$$p(Y_{ijk}|\Theta_{ij}, e_k^+, e_k^-) = (1 - I_{ijk}^Y) + I_{ijk}^Y((1 - \Theta_{ij})(Y_{ijk}e_k^+ + (1 - Y_{ijk})(1 - e_k^+)) + \Theta_{ij}(Y_{ijk}(1 - e_k^-) + (1 - Y_{ijk})e_k^-)) \tag{42}$$

Under our standard assumption of conditional independence, the joint likelihood of the data is simply the product of the individual arc likelihoods:

$$p(Y|\Theta, e^+, e^-) = \prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e_k^+, e_k^-) \tag{43}$$

which can be seen quite readily to be a simple generalization of Eq. (11) to the multiple observer case.

2.4.3. Computation of the posterior

Having determined the joint likelihood of the data, we are now in a position to write down the posterior. Using Eq. (43) and Bayes' theorem:

$$p(\Theta, e^+, e^-|Y) \propto p(Y|\Theta, e^+, e^-)p(\Theta)p(e^+)p(e^-) \tag{44}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e_k^+, e_k^-)\right)\left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\right)\left(\prod_{i=1}^{M}\text{Beta}(e_i^+|\alpha_i^+, \beta_i^+)\right) \times \left(\prod_{i=1}^{M}\text{Beta}(e_i^-|\alpha_i^-, \beta_i^-)\right) \tag{45}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e_k^+, e_k^-)\right)\left(\prod_{i=1}^{M}\text{Beta}(e_i^+|\alpha_i^+, \beta_i^+)\right) \times \left(\prod_{i=1}^{M}\text{Beta}(e_i^-|\alpha_i^-, \beta_i^-)\right) \tag{46}$$

Analytically, the joint posterior given by Eq. (44) is somewhat difficult to work with. However, as in the case of the single observer model, we can exploit the factorability of the posterior to easily derive the full conditionals of the model, which in turn happen to be appropriate for the straightforward implementation of a Gibbs sampler. In the case of the criterion graph, for instance, Eq. (46) already demonstrates the factorization needed to

derive the conditional distribution:

$$p(\Theta|e^+, e^-, Y) = \prod_{i=1}^{N}\prod_{j=1}^{N}\left[(1 - I_{ij}^{\Theta}) + I_{ij}^{\Theta} \times B\left(\Theta_{ij}\left|\frac{\Phi_{ij}\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|1 - e_k^-)]}{\Phi_{ij}\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|1 - e_k^-)] + (1 - \Phi_{ij})\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|e_k^+)]}\right.\right)\right] \tag{47}$$

Note that the presence of multiple error parameters does not interfere with the conditional independence of arcs. Hence, we can decompose Eq. (47) to provide the conditional distribution for any given arc subject to $I_{ij}^{\Theta} = 1$:

$$\Theta_{ij}|e^+, e^-, Y \sim B\left(\Theta_{ij}\left|\frac{\Phi_{ij}\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|1 - e_k^-)]}{\Phi_{ij}\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|1 - e_k^-)] + (1 - \Phi_{ij})\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}B(Y_{ijk}|e_k^+)]}\right.\right) \tag{48}$$

To find the full conditionals for the various error parameters, we follow a very similar procedure to that pursued for the single observer model; the difference in this case is that we must decompose the posterior into separate factors for each of the $2M$ parameters.

$$p(\Theta, e^+, e^-|Y) \propto \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e_k^+, e_k^-)\right)\left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij})\right) \times \left(\prod_{i=1}^{M}\mathrm{Beta}(e_i^+|\alpha_i^+, \beta_i^+)\right)\left(\prod_{i=1}^{M}\mathrm{Beta}(e_i^-|\alpha_i^-, \beta_i^-)\right) \tag{49}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M}[(1 - I_{ijk}^{Y}) + I_{ijk}^{Y}(\Theta_{ij}B(Y_{ijk}|1 - e_k^-) + (1 - \Theta_{ij})B(Y_{ijk}|e_k^+))]\right) \times \left(\prod_{i=1}^{M}\mathrm{Beta}(e_i^-|\alpha_i^-, \beta_i^-)\right)\left(\prod_{i=1}^{M}\mathrm{Beta}(e_i^+|\alpha_i^+, \beta_i^+)\right) \tag{50}$$

$$= \left[\prod_{k=1}^{M} \left(\prod_{(i,j):\Theta_{ij}=1, I^Y_{ijk}=1} B(Y_{ijk}|1-e^-_k) \right) \mathrm{Beta}(e^-_i|\alpha^-_i, \beta^-_i) \right]$$
$$\times \left[\prod_{k=1}^{M} \left(\prod_{(i,j):\Theta_{ij}=0, I^Y_{ijk}=1} B(Y_{ijk}|e^+_k) \right) \mathrm{Beta}(e^+_i|\alpha^+_i, \beta^+_i) \right] \left(\prod_{i=1}^{N}\prod_{j=1}^{N} p(\Theta_{ij}) \right) \quad (51)$$

Conditioning on Θ, the network prior is now a constant, and each error parameter is conditionally independent. This allows us, as before, to accumulate correct and incorrect edge observations by parameter. For instance, the conditional for an arbitrary e^+_k is as follows:

$$p(e^+_k|e^+_{-k}, e^-, \Theta, Y) \propto \left[\prod_{(i,j):\Theta_{ij}=0, I^Y_{ijk}=1} B(Y_{ijk}|e^+_k) \right] \mathrm{Beta}(e^+_k|\alpha^+_k, \beta^+_k) \quad (52)$$

$$\propto \left[(e^+_k)^{\sum_{i=1}^N \sum_{j=1}^N I^Y_{ijk}(1-\Theta_{ij})Y_{ijk}} (1-e^+_k)^{\sum_{i=1}^N \sum_{j=1}^N I^Y_{ijk}(1-\Theta_{ij})(1-Y_{ijk})} \right]$$
$$\times ((e^+_k)^{\alpha^+_k - 1}(1-e^+_k)^{\beta^+_k - 1}) \quad (53)$$

$$= (e^+_k)^{\alpha^+_k + \left(\sum_{i=1}^N \sum_{j=1}^N I^Y_{ijk}(1-\Theta_{ij})Y_{ijk}\right) - 1} (1-e^+_k)^{\beta^+_k + \left(\sum_{i=1}^N \sum_{j=1}^N I^Y_{ijk}(1-\Theta_{ij})(1-Y_{ijk})\right) - 1} \quad (54)$$

where e^+_{-k} is the vector of all false positive parameters other than e^+_k. (The conditional independence of the error parameters implies that explicit conditioning on the other false positive (or false negative) error parameters is not actually required, but the notation is retained for clarity.) Once again, we observe that the unnormalized conditional distribution for each false positive parameter is proportional to the Beta density, which gives us the exact conditional:

$$e^+_k|e^+_{-k}, e^-, \Theta, Y$$
$$\sim \mathrm{Beta}\left(\alpha^+_k + \sum_{i=1}^{N}\sum_{j=1}^{N} I^Y_{ijk}(1-\Theta_{ij})Y_{ijk}, \beta^+_k + \sum_{i=1}^{N}\sum_{j=1}^{N} I^Y_{ijk}(1-\Theta_{ij})(1-Y_{ijk}) \right) \quad (55)$$

For the false negative parameters, we factor the conditional in a similar way:

$$p(e^-_k|e^-_{-k}, e^+, \Theta, Y) \propto \left[\prod_{(i,j):\Theta_{ij}=1, I^Y_{ijk}=1} B(Y_{ijk}|1-e^-_k) \right] \mathrm{Beta}(e^-_k|\alpha^-_k, \beta^-_k) \quad (56)$$

$$\propto \left[(e^-_k)^{\sum_{i=1}^N \sum_{j=1}^N I^Y_{ijk}\Theta_{ij}(1-Y_{ijk})} (1-e^-_k)^{\sum_{i=1}^N \sum_{j=1}^N I^Y_{ijk}\Theta_{ij}Y_{ijk}} \right]$$
$$\times (e^-_k)^{\alpha^-_k - 1}(1-e^-_k)^{\beta^-_k - 1} \quad (57)$$

$$= (e_k^-)^{\alpha_k^- + \left(\sum_{i=1}^N \sum_{j=1}^N I_{ijk}^Y \Theta_{ij}(1-Y_{ijk})\right)-1} (1-e_k^-)^{\beta_k^- + \left(\sum_{i=1}^N \sum_{j=1}^N I_{ijk}^Y \Theta_{ij} Y_{ijk}\right)-1} \quad (58)$$

leading to the Beta distribution

$$e_k^- | e_{-k}^-, e^+, \Theta, Y \sim \text{Beta}\left(\alpha_k^- + \sum_{i=1}^N \sum_{j=1}^N I_{ijk}^Y \Theta_{ij}(1-Y_{ijk}), \beta_k^- + \sum_{i=1}^N \sum_{j=1}^N I_{ijk}^Y \Theta_{ij} Y_{ijk}\right) \quad (59)$$

This completes the full conditionals for the multiple observer model. To implement the Gibbs sampler, we follow the same general procedure as was employed for the single observer model, with the slight modification of separate draws for each error parameter. One potential implementation of such a scheme is the following:

1. **procedure** Draw from $\Theta, e^+, e^- | Y$ (Multiple Observer Model)
2. Draw $\Theta^{(1)}$ from $p(\Theta)$
3. **for** $j \in (1, \ldots, M)$ **do**
4. Draw $e_j^{+(1)}$ from $p(e_j^+)$
5. Draw $e_j^{-(1)}$ from $p(e_j^-)$
6. **end for**
7. $i := 2$
8. **repeat**
9. Draw $\Theta^{(i)}$ from $p(\Theta | e^{+(i-1)}, e^{-(i-1)}, Y)$
10. **for** $j \in (1, \ldots, M)$ **do**
11. Draw $e_j^{+(i)}$ from $p(e_j^+ | \Theta^{(i)}, e^{-(i-1)}, Y)$
12. Draw $e_j^{-(i)}$ from $p(e_j^- | \Theta^{(i)}, e^{+(i-1)}, Y)$
13. **end for**
14. $i := i + 1$
15. **until** $\Theta^{(\cdot)}, e^{+(\cdot)}, e^{-(\cdot)} \sim \Theta, e^+, e^- | Y$
16. **return** $\Theta^{(\cdot)}, e^{+(\cdot)}, e^{-(\cdot)}$

Note that the usual provisos regarding chain convergence apply here; in particular, a longer burn-in time may be required due to the increased number of dimensions for the multiple observer case.

2.5. Multiple error probabilities with exponential random graph priors

Although we have focused thusfar exclusively on the case of Bernoulli graph priors, the error models shown here can be used under arbitrary exponential random graph priors as well. The price for this generality is a certain loss of analytical and computational simplicity; nevertheless, we shall here outline the basic results needed to use the multiple observer model with such a prior.

2.5.1. Assignment of priors

Error priors in this case are unchanged from our earlier treatment, and we continue to assume that error rates are a priori independent of the criterion graph. As for the network

prior, we now assume that it takes the form

$$p(\Theta) = \begin{cases} \dfrac{e^{\sum_{i=1}^{\ell}\phi_i t_i(\Theta)}}{\sum_{\Theta' \in Q} e^{\sum_{i=1}^{\ell}\phi_i t_i(\Theta)}} & \text{if } \Theta \in Q \\ 0 & \text{otherwise} \end{cases} \tag{60}$$

$$= I^{Q}(\Theta)\frac{e^{\sum_{i=1}^{\ell}\phi_i t_i(\Theta)}}{\kappa(\phi, t)} \tag{61}$$

where $\phi \in \mathbb{R}^{\ell}$, t is a vector of real-valued functions of Θ (and, implicitly, I^{Θ}), Q is the set of all realizable adjacency matrices for G, and I^{Q} is given by

$$I^{Q} = \begin{cases} 1 & \text{if } \Theta \in Q \\ 0 & \text{otherwise} \end{cases} \tag{62}$$

Eq. (60) is the exponential random graph (ERG) distribution (also called the p^* model (Wasserman and Pattison, 1996)), a general form which takes the Bernoulli graph distribution as a special case. By appropriate choice of ϕ and t, one can parameterize any Q-conditioned graph distribution in this manner; thus, for our purposes, this distribution serves as a fully generic prior family for the criterion graph.

The conventional difficulty in working with Eq. (60) lies in the calculation of the normalizing factor, $\kappa(\phi, t)$. Unlike the normalizing constants with which we have been working, κ varies with the choice of ϕ and t, making inference for these parameters difficult. (See Anderson et al. (1999b) for a very accessible introduction to estimation of exponential random graph models using pseudo-likelihood methods, and Snijders (2002) for more recent work using a Monte Carlo Maximum Likelihood (MCML) approach.) Fortunately, this particular problem does not present an obstacle to us, as our inferences are on Θ conditional on ϕ and t rather than on ϕ conditional on Θ and t. For our purposes, then, we will be able to use the unnormalized exponential random graph density, relying on MCMC simulation to carry out implicit normalization.

A secondary difficulty with the ERG distribution is that of selecting values of ϕ and t which adequately represent our prior knowledge regarding the criterion graph. Unlike the "overdispersed" case of the Bernoulli prior, arbitrary ERG distributions are often highly concentrated (Snijders, 2002) and their properties are not completely understood at this time. Appropriate use of a non-trivial ERG prior may considerably improve inference, but *inappropriate* use of the same may substantially degrade it; researchers are (as always) encouraged to select ERG hyperparameters carefully, and to err on the side of more diffuse models where available. Subject to this caveat, an obvious source for reasonable prior information would be ERG estimates for other, similar networks having the same size as G. Provided that a diffuse model can be chosen within the range of prior estimates, such a model would be expected to form a reasonable a priori distribution for the criterion graph. Having chosen an ERG prior for Θ, we select error rate priors as per the multiple observer model treated above. Under the prior independence assumption, changing the network prior does not impact our initial beliefs regarding error rates, and, indeed, much of this component of the analysis will carry forward into the present case.

2.5.2. Assumed likelihood

Like the error priors, the likelihood for this model is unchanged from the multiple observer model with Bernoulli network priors. Specifically, we continue to employ Eqs. (42) and (43) to represent the data generation process.

2.5.3. Computation of the posterior

Following the familiar drill, we derive the joint posterior for the multiple observer model with exponential random graph priors:

$$p(\Theta, e^+, e^-|Y) \propto p(Y|\Theta, e^+, e^-)p(\Theta)p(e^+)p(e^-) \tag{63}$$

$$= \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e_k^+, e_k^-)\right) p(\Theta) \left(\prod_{i=1}^{M} \text{Beta}(e_i^+|\alpha_i^+, \beta_i^+)\right) \times \left(\prod_{i=1}^{M} \text{Beta}(e_i^-|\alpha_i^-, \beta_i^-)\right) \tag{64}$$

Factorization of the posterior for the error parameters can be performed precisely as per the earlier multiple observer model, and Eqs. (55) and (59) can be used to take posterior conditional draws for the false positive and false negative error parameters (respectively). We cannot, however, factor the posterior in such a way as to allow us to draw easily from the conditional of Θ, as was possible under the Bernoulli prior. The unnormalized density with which we are left is

$$p(\Theta|e^+, e^-, Y) \propto \left(\prod_{i=1}^{N}\prod_{j=1}^{N}\prod_{k=1}^{M} p(Y_{ijk}|\Theta_{ij}, e_k^+, e_k^-)\right) \left(I^Q(\Theta) e^{\sum_{i=1}^{\ell}\phi_i t_i(\Theta)}\right) \tag{65}$$

which, for reference, we name explicitly as

$$= q(\Theta|e^+, e^-, Y) \tag{66}$$

While we cannot sample directly from this density, we may nevertheless use it as part of an overall sampling scheme to take posterior draws. In particular, we may employ a Metropolis–Hastings algorithm which combines a "Gibbs step" to sample error parameters with a "Metropolis step" to sample Θ. While the Gibbs step always accepts new proposals (since it draws directly from the full conditionals), the Metropolis step first makes a proposal based on a perturbation of the current graph and then accepts the proposal with a probability equal to the ratio of the unnormalized conditional densities of the proposed and current graphs (respectively). Such an algorithm can be implemented as follows:

1. **procedure** Draw from $\Theta, e^+, e^-|Y$ (ERG Network Prior)
2. Draw $\Theta^{(1)}$ from $p(\Theta)$
3. **for** $j \in (1, \ldots, M)$ **do**
4. Draw $e_j^{+(1)}$ from $p(e_j^+)$
5. Draw $e_j^{-(1)}$ from $p(e_j^-)$
6. **end for**

$i := 2$
repeat
  for $j \in (1, \ldots, N)$ **do**
    for $k \in (1, \ldots, N)$ **do**
      if $I^{\Theta}_{jk} = 1$ **then**
        Draw $\Theta^{(i)}_{jk} \sim B((1-\tau)\Theta^{(i-1)}_{jk} + \tau(1-\Theta^{(i-1)}_{jk}))$
      end if
    end for
  end for
Draw $u \sim \text{Unif}(0, 1)$
if $u > (q(\Theta^{(i)}|e^+, e^-, Y))/(q(\Theta^{(i-1)}|e^+, e^-, Y))$ **then**
  $\Theta^{(i)} := \Theta^{(i-1)}$
end if
for $j \in (1, \ldots, M)$ **do**
  Draw $e^{+(i)}_j$ from $p(e^+_j|\Theta^{(i)}, e^{-(i-1)}, Y)$
  Draw $e^{-(i)}_j$ from $p(e^-_j|\Theta^{(i)}, e^{+(i-1)}, Y)$
  end for
  $i := i + 1$
until $\Theta^{(\cdot)}, e^{+(\cdot)}, e^{-(\cdot)} \sim \Theta, e^+, e^-|Y$
return $\Theta^{(\cdot)}, e^{+(\cdot)}, e^{-(\cdot)}$

Note that τ is assumed to be chosen from the (0, 1) interval; higher values of τ reflect larger "jumps" in the space of graphs, but can also result in the chain's becoming "stuck" in a particular configuration. In general, experimentation will be required to identify values of τ which provide optimal convergence for a given model. It is important to note, too, that the particular algorithm shown above is not the only MCMC which can be used to generate posterior draws for this model: in addition to alternative jumping rules, Metropolis–Hastings can be employed jointly for all parameters, rather than merely a subset as shown here. Identification of optimal algorithms for drawing from error models with exponential random graph priors would seem to be an important area for future research.

2.6. Estimating network variables from posterior distributions

We have now developed four different Bayesian models of the network inference/informant accuracy problem, and have shown in each case how we may simulate draws from the relevant posterior distribution. For certain applications (e.g. estimating the accuracy of particular actors) this may be sufficient; in general, however, our primary interest will be in various quantities which are derived from the criterion graph itself. Given that we are uncertain about the criterion graph, these quantities will necessarily be random variables, with distributions which depend upon the posterior of the criterion. In the following section, then, we shall demonstrate the application of the network posterior to two standard inference problems: point estimation of the criterion graph, and inference regarding graph and node level indices for the criterion graph.

To illustrate these procedures, we will utilize posterior draws from a CSS of Krackhardt (1987a) under the multiple observer model. The data set in question contains reported friendship and advice-seeking relations among 21 management personnel within a high-tech firm, and is a "classic" CSS study. For purposes of illustration, posterior draws were taken independently for both relations. The network prior for this analysis was chosen such that $\Phi_{ij} = 0.3$ for all arcs, and all individual error parameters were given Beta(3, 5) priors. Posterior draws were taken using a Gibbs sampler, with three Markov chains being employed, each having a burn-in of 500 iterations. (Chain convergence for each scalar estimand was also assessed using the Gelman and Rubin (1992) $\sqrt{\hat{R}}$ measure.) After burn-in, 500 draws were taken from each chain.[15] The 1500 posterior draws derived from this process were then used in the analyses below.

2.6.1. Network estimators

Given the joint posterior, a problem of obvious interest is that of estimating the criterion graph. Numerous estimators of Θ are possible, two being the joint posterior mode and the (marginal) edgewise posterior mode. The first of these is obtained by numerical maximization of $p(\Theta, e^+, e^-|Y)$; given the discrete nature and high dimensionality of the parameter space, heuristic optimization (e.g. simulated annealing, genetic algorithms) will generally be required to identify this estimator.

The second estimator—the edgewise posterior mode—differs from the first in that it seeks modal states for each arc of the criterion graph, rather than the graph as a whole. Formally:

$$\hat{\Theta}_{ij} = \begin{cases} 1 & \text{if } p(\Theta_{ij}|Y) \geq 0.5 \\ 0 & \text{if } p(\Theta_{ij}|Y) < 0.5 \end{cases} \tag{67}$$

and the corresponding $\hat{\Theta}$ is simply the collection of the individual arc estimators. This estimator has the advantage of being quite simple to obtain, being well-approximated by the dichotomized graph mean of a series of posterior draws. The principal disadvantage of this estimator is that it averages across any dependence between arcs, and hence may be misleading in some cases. For our illustrative data set, the adjacency matrix of the edgewise modal graph is shown in Table 1. A visualization of the edgewise modal graph for the Krackhardt advice network is shown in Fig. 1.

It should be noted that both modes described above are point estimates, and as such necessarily discard much of the full information of the posterior distribution. For many applications, then, it may be more prudent to use the posterior distribution, rather than a graph estimator. Nevertheless, when it is desirable to have some particular estimate of the criterion structure, these may often be reasonable options. For special cases in which an estimator is required which meets particular criteria (e.g. minimizing some convex function of differences in graph statistics vis a vis the true criterion), it may also be desirable to identify a particular graph which minimizes a posterior predictive loss function. For a given loss, L, such an estimator $\hat{\Theta}_L$ can be obtained by numerical minimization of

$$L(\hat{\Theta}_L, \Theta^*), \quad \Theta^* \sim \Theta|Y \tag{68}$$

[15] *S* code (written for the *R* statistical computing system) to fit this model is available from the author.

Table 1
Edgewise modal graph for advice (adjacency matrix)

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21
a1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0
a2	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1
a3	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0
a4	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1
a5	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	1	0
a6	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1
a7	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1
a8	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1
a9	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0
a10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0
a11	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0
a12	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
a13	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	1	1	0	0
a14	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0	1	1	1	1
a15	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	0
a16	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
a17	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
a18	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	1
a19	0	1	1	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0
a20	0	1	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0
a21	0	1	0	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0

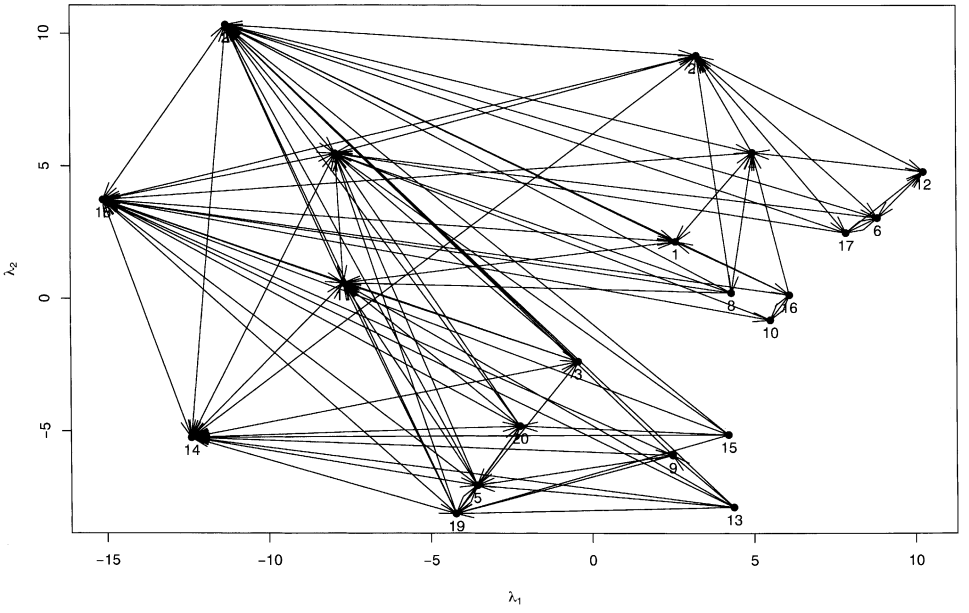

Fig. 1. Edgewise modal graph, advice network.

Note that the posterior mode minimizes the 0–1 loss, and its edgewise counterpart minimizes absolute errors (summed across arcs). Investigation of more complex losses (e.g. those motivated by applied problems) is an important avenue for further study.

2.6.2. Graph and node level indices

The vast majority of classical network analysis is founded on substantively important secondary indices such as degree, betweenness, and closeness (at the nodal level) and centralization, hierarchy, and connectedness (at the graph level). Many of these indices are highly sensitive to minor changes in network structure; yet, traditional network analysis has generally examined these indices under the assumption of error-free data. One useful application of the network posterior, then, is to permit estimation of graph and node level indices in the presence of measurement error. The quantification of uncertainty, in particular, is useful here: even where the posterior is diffuse, knowledge of this fact may illuminate subsequent analysis. This last is particularly true of *comparisons* of network indices, where changes in the uncertainty associated with point estimates may lead to substantively distinct conclusions.

By way of illustration, then, we here provide summaries of posterior intervals for Freeman degree centralization and nodal degree for the Krackhardt advice network (see Table 2). Note how, even for a simple measure such as degree, quantification of the uncertainty associated with the criterion graph can clearly affect posterior inference. In addition to the simple

Table 2
Posterior predictive distribution of Freeman degree, advice network

Index	Minimum	1st *Q*	Median	Mean	3rd *Q*	Maximum
Degree, a1	7	9	10	10.21	11	14
Degree, a2	23	24	25	25.05	26	29
Degree, a2	6	8	8	8.18	9	12
Degree, a4	8	10	10	9.88	10	11
Degree, a5	9	12	12	12.28	13	14
Degree, a6	6	7	8	7.87	8	10
Degree, a7	16	19	20	19.61	20	23
Degree, a8	5	7	8	7.94	8	12
Degree, a9	6	7	7	7.34	8	10
Degree, a10	3	5	6	6.15	7	12
Degree, a11	12	14	15	14.72	15	17
Degree, a12	4	6	7	6.57	7	9
Degree, a13	7	7	8	7.99	8	11
Degree, a14	19	21	22	21.99	23	25
Degree, a15	5	6	7	7.05	8	10
Degree, a16	5	7	7	7.23	8	10
Degree, a17	5	6	7	6.8	7	9
Degree, a18	22	24	24	24.13	24	27
Degree, a19	10	13	13	13.01	13	16
Degree, a20	6	11	11	11.45	12	16
Degree, a21	13	15	15	15.13	16	17
Centralization	0.31	0.35	0.36	0.37	0.38	0.45

summaries shown here, it is fairly trivial to compute quantities such as, for instance, the posterior probability that actor 5 has a higher Freeman degree than actor 4 (in this case, the probability in question is approximately 0.985). Such an approach is clearly more powerful than classical methods (which can only examine relative likelihoods), and is obviously more robust than traditional network techniques which treat the data under examination as error-free. The Bayesian modeling approach, then, is useful not only because of the ease with which it allows us to construct theoretically motivated models, but also because of the inferential uses to which the output of those models can be put.

2.6.3. Estimated informant accuracy

While we may often be interested solely in the criterion graph (in which case we estimate individual accuracy parameters only because such are necessary for the previous problem), this is not always the case; in some cases, it is the accuracy of individuals which is of interest (e.g. Krackhardt (1990)). Given a series of draws from the posterior distribution, it is quite trivial to examine the posterior distribution of individual error parameters, and thereby to gain a sense of actors' abilities to perceive their social surroundings. These distributions may in turn be used to create point estimates which may be employed in subsequent analyses, though as always it is preferable to employ posterior draws for this purpose where possible so as to avoid losing the distributional information contained therein.

For our advice network, then, the estimated posterior quantiles for each parameter are given in Tables 3 and 4. Note that while the above represent quantiles of the posterior marginals, it is important to bear in mind that the error parameters are not independent in the

Table 3
Posterior marginal probability of false positives (e^+) by observe, advice network

Observer	Minimum	1st *Q*	Median	Mean	3rd *Q*	Maximum
a1	0.45	0.54	0.56	0.55	0.57	0.66
a2	0.05	0.1	0.12	0.12	0.13	0.19
a3	0.22	0.3	0.32	0.32	0.34	0.41
a4	0.21	0.29	0.31	0.31	0.33	0.41
a5	0.19	0.27	0.29	0.29	0.31	0.38
a6	0	0.02	0.02	0.03	0.03	0.07
a7	0.15	0.2	0.22	0.22	0.24	0.3
a8	0.06	0.1	0.12	0.12	0.13	0.18
a9	0.15	0.21	0.22	0.22	0.24	0.3
a10	0.12	0.18	0.19	0.19	0.21	0.28
a11	0.02	0.05	0.06	0.06	0.07	0.12
a12	0.08	0.14	0.15	0.15	0.17	0.28
a13	0	0.02	0.02	0.02	0.03	0.06
a14	0.14	0.2	0.21	0.21	0.23	0.3
a15	0.06	0.11	0.12	0.12	0.13	0.21
a16	0	0.01	0.01	0.01	0.02	0.06
a17	0	0.02	0.02	0.02	0.03	0.05
a18	0.08	0.14	0.16	0.16	0.17	0.24
a19	0.02	0.05	0.06	0.06	0.07	0.12
a20	0.05	0.07	0.08	0.09	0.1	0.17
a21	0.21	0.28	0.3	0.3	0.32	0.38

Table 4
Posterior marginal probability of false negatives (e^-) by observer, advice network

Observer	Minimum	1st *Q*	Median	Mean	3rd *Q*	Maximum
a1	0.05	0.1	0.12	0.12	0.14	0.23
a2	0.24	0.35	0.38	0.38	0.41	0.55
a3	0.13	0.21	0.23	0.23	0.26	0.37
a4	0.08	0.14	0.16	0.16	0.18	0.32
a5	0.08	0.15	0.17	0.17	0.2	0.29
a6	0.47	0.58	0.61	0.61	0.64	0.74
a7	0.14	0.23	0.26	0.26	0.29	0.4
a8	0.08	0.17	0.19	0.2	0.22	0.33
a6	0.17	0.26	0.29	0.29	0.32	0.42
a10	0.13	0.22	0.25	0.25	0.28	0.37
a11	0.36	0.47	0.5	0.5	0.53	0.69
a12	0.17	0.28	0.3	0.3	0.33	0.43
a13	0.48	0.59	0.62	0.62	0.66	0.76
a14	0.24	0.36	0.39	0.39	0.42	0.52
a15	0.3	0.43	0.46	0.46	0.49	0.6
a16	0.46	0.57	0.6	0.6	0.63	0.72
a17	0.5	0.64	0.67	0.67	0.7	0.79
a18	0.11	0.19	0.21	0.21	0.24	0.34
a19	0.17	0.27	0.3	0.3	0.33	0.44
a20	0.35	0.46	0.49	0.49	0.52	0.62
a21	0.04	0.11	0.13	0.13	0.15	0.26

posterior. Thus, contrary to what one might at first blush expect, we find that the probability of any actor within the advice network having error parameters leading to perverse inferences is vanishingly small (indeed, this was not observed for any actor for any draws in the advice network, and was observed on only 3 out of 31,500 possible occasions (and then only weakly so) for draws concerning the friendship network). Similarly, the fact that the probability of false negatives exceeds that of false positives in approximately 74% of cases in advice network draws (and in over 99% of cases in friendship network draws) is an important and interesting observation which follows from direct examination of the draws themselves. This suggests, for instance, a general asymmetry between positive and negative errors: quite possibly, informants are reasonably accurate when reporting ties, but generally quite inaccurate when reporting holes. Such a finding is consistent with a model which assumes that informant inaccuracy is largely the result of limited information, and that ties are not uniformly visible; if this is the case, we should expect to see an even stronger asymmetry with private relationships (such as friendship) than with public ones.[16] Using the present model, then, we can not only gain general estimates of informant accuracy, but also ask broader questions regarding the types of mistakes made by actors, the informativeness of individual responses, and the like; these questions are easily assessed by examination of posterior draws.

[16] In fact, precisely this relationship is found in the present data set, with the median per-draw difference in error scores being 0.14 for advice and 0.56 for friendship.

3. Discussion

We have, in previous sections of this paper, outlined a Bayesian modeling approach to (some aspects of) the problem of informant accuracy in social network analysis. In the course of this development, a number of deeper issues have arisen which, while important to the progress of our research, are somewhat tangential to the specific focus of the present work. Despite this, it behooves us to give at least some consideration to three particularly important issues: the robustness of the modeling approach to structurally-correlated errors; the problem of improving data collection for informant self-reports in social network research; and the concept validity of the criterion graph (on which the present enterprise obviously depends). This section, then, will initiate a discussion of these four matters which, if insufficient to resolve them, will hopefully serve to highlight them for subsequent research.

3.1. Robustness to structurally correlated accuracy parameters

A common finding of a variety of network accuracy studies (e.g. Romney and Faust, 1982; Freeman et al., 1987; Krackhardt, 1990; Freeman, 1992; Krackhardt and Kilduff, 1999) is that network position is differentially associated with informant accuracy. While the specific relationships involved are only partially understood at present,[17] the existence (or at least plausibility) of a distance/accuracy effect is generally agreed upon by network analysts. Given this, it is important to note that the models presented here do not directly incorporate such information into the joint prior distribution of Θ, e^+, and e^-. This raises a question of robustness: if real-world data were, in fact, to contain structurally correlated error rates, to what extent would this degrade model performance? Clearly, our simple models cannot be *optimal* in such circumstances, as we omit the possibility of allowing knowledge regarding structure/error associations to inform our inferences. Nevertheless, the models may be *robust* in the sense that their accuracy is not diminished substantially in the presence of these dependencies. Before we seek to deploy these models in a real-world context, it behooves us to assess this possibility.

As a preliminary measure, Monte Carlo tests of these models were conducted for small ($N = M = 15$) CSS data sets with centrality-correlated error rates. Overall, the basic finding of this robustness testing was that even when informants' accuracy parameters were *perfectly* (inversely) correlated with criterion centrality,[18] there was little degradation in model performance. Estimates of individual accuracy parameters, in particular, did not appear to suffer under structurally correlated accuracy, and errors in estimation were similarly distributed in uniform and correlated cases. Estimates of the criterion graph did show some increase in error, but this increase was small (affecting only 1% of arcs on average) and similarly distributed across different correlation models. Given that we would expect the real-world relationship between centrality and accuracy to be weaker than those tested, this

[17] It has been argued variously that centrality, transitivity, and social distance (among other factors) may affect accuracy in a variety of ways; unfortunately, these inferences are based on a small number of data sets and, in some cases, the criterion graph has been inferred via ad hoc methods such as the Locally Aggregated Structure (Krackhardt, 1987a).

[18] A repetition of this experiment with correlations reversed (high centrality leading to low accuracy) produced the same qualitative results.

test allows us to be at least marginally confident that such factors will not interfere with the use of the informant accuracy/network inference models thusfar presented. Of course, this simple test does not demonstrate robustness to all potential structure/accuracy relationships, but it does suggest that one of the broadest and most commonly hypothesized associations is unlikely to seriously impair model performance.

3.2. Suggestions for improved data collection

In the introduction to this paper, four general problems relating to the informant accuracy in network research were mentioned. As indicated, we have focused primarily on the last, namely the development of inferential techniques for quantifying (and hopefully reducing) the uncertainty inherent in this form of data. This endeavor, however, is strongly related to another: the development and deployment of data collection strategies which facilitate the reduction of uncertainty regarding quantities of interest. While a variety of issues are involved in this pursuit, we shall here constrain ourselves to two matters, namely sampling strategies employed in eliciting network data from informants and the handling of missing data.

3.2.1. Sampling strategies

The standard procedure for eliciting informant reports regarding social structure is generally to provide each member of the social network with a survey instrument which elicits his or her ties to others within the group. In some cases, this is extended by link-tracing (Thompson and Frank, 2000), inquiring into the existence of incoming ties from non-group members, or (as in ego net research) asking for ties among adjacent alters; nevertheless, these approaches are less uniformly deployed than the first. The sub-optimality of this standard procedure from the point of view of network inference can easily be appreciated by counting the number of repeated observations accorded each arc: plainly, the standard procedure counts each arc but once, and even extending this to incoming ties provides only two observations. With so few observations, it is hardly surprising that our ability to infer social structure is so problematic! Given even minor deviations from perfect reporting, procedures which supply only one to two replications per arc are unlikely to provide sufficient data for reasonable inference on the criterion graph.

One alternative to this procedure is the elicitation of cognitive social structures. Though developed explicitly as a cognitivist tool, the CSS instrument is highly desirable from a classical perspective due to the fact that it provides a reasonable number of repeated observations—N, to be precise—on each arc. Further, the fact that the CSS elicits observations from all network members means that it can be considered to be an ignorable design so long as the complete data set is defined in terms of all participant observations. CSS data, then, is of much greater potential value to the network analyst than traditional data, particularly when employed in conjunction with inferential tools which allow for inference across arcs and across actors.

For all its benefits, there is a clear drawback to the CSS design: due to the fact that each informant is asked to report on all arcs, the number of items on a CSS instrument increases on the order of N^2. This polynomial growth stands in sharp contrast to the linear growth of instrument complexity for traditional approaches, and severely limits the size of networks which can be examined in this fashion. A 50 node network, for instance, requires each

subject to consider 2500 items for every relation examined; plainly this stretches the limits of informant endurance. Unfortunately, we are often interested in networks which are of even larger sizes, which all but eliminates the CSS from use in a wide range of settings.

Given the above problems, it may be sensible to consider alternatives to both the traditional and CSS data collection strategies, particularly for large networks. Such an alternative should be ignorable, should provide multiple observations on each arc, and should provide multiple observations on each informant, while maintaining linear complexity in network size. One data collection strategy which fulfills these requirements is what we shall here call a *K-replication balanced arc sampling design*; while we will not consider all of its properties in detail, we shall outline the basic procedure by which it may be employed.

The core intuition of the *K*-replication balanced arc sampling design is that if one desires to have *K* observations on each arc in a directed graph, one need only ask each informant to supply *KN* observations. (This follows from the fact that *N* subjects reporting *KN* arcs results in KN^2 observations, enough to allow *K* per arc.) The challenge, then, is to allocate the arcs sampled in such a way as to maintain ignorability. One simple means of doing so is to randomly allocate arcs to instruments such that (A) each informant is given exactly *KN* arcs on which to report, and (B) each arc is reported on exactly *K* times. Such a design is then ignorable, as observations not included are missing at random (see Gelman et al. (1995)), and balanced (as each informant and arc contribute equally to the joint likelihood). With the *K*-replication balanced arc sampling design, one can gain many of the advantages of a CSS design (albeit on a more limited basis) without incurring the same complexity penalty. Inferential methods such as those discussed here can then be employed to estimate the criterion graph, which can subsequently be used for classical network analysis purposes.

3.2.2. Elicitation and missing data

If, on the one hand, standard practice in network research prompts us to collect too little data on ties between actors, one might also argue that it occasionally prompts us to collect too *much*: because it is customary (even in CSS research) to obtain informants' assessments of ties in forced-choice framework, we generally have no means of separating out informant assessments based on strong personal information from those made in near-total ignorance. This is a consequential omission. If, for instance, informants knowing nothing of the interactions between two actors A and B generally respond to a question of the form "does A go to B for help or advice at work?" with a "no" response, then we may expect to find very high rates of false negatives in populations where actors are not well-informed regarding each others' behaviors. Such a conclusion may be very misleading: it may well be that informants who are confident that A does not go to B for help or advice (i.e. those who would choose a "no" option if given a "do not know" alternative) are quite accurate, but that we cannot discern this fact due to the large number of informants who are using "no" as a default option. Allowing informants to indicate their ignorance thus has the potential to substantially improve our assessments of informant accuracy, and by extension to draw stronger network inferences.

This insight regarding the difficulties of forced-choice questions for judgment tasks in which informant information varies is hardly novel; results linking inappropriate use of forced choice to measurement problems are well-established in the methodological literature (Schwarz, 1999; Schwarz et al., 1998) and mention of it can be found even in introductory

texts (e.g. DeVellis (1991); Bailey (1987)). Given this, why have network analysts continued to employ forced choice when eliciting judgments from informants? A likely possibility is simply that there have been, historically, very few alternatives.[19] Without methods for dealing with missing data in network research, "do not know" responses would likely be coded as tie-missing,[20] thus removing much of the potential benefit of departing from forced-choice in the first place. A potential side benefit, then, of the models presented here is that (by allowing inferences in the presence of missing data) they facilitate the use of elicitation methods which do not force informants to provide responses for ties of which they are ignorant.

Another side benefit of the ability to draw inferences with missing data lies in robustness testing: by expanding the data set to include hypothetical actors whose ties are unobserved (missing), but for whose ties we may still have prior knowledge, we can get some sense of the uncertainty associated with accidentally omitting members of the network. By drawing from the combined network posterior, we can compare probability intervals for graph and node level indices (for instance) with those derived from draws from the initial (non-missing) posterior; if our initial conclusions continue to hold, then we may reasonably assume that they would likely be unaltered by the addition of the hypothetical actors. Clearly, certain assumptions must be made in such an analysis (e.g. unobserved actors must be ignorably missing, we must have network priors for the unobserved arcs), but the approach does provide a simple initial robustness test and can be performed using existing tools. Like the implications of the Bayesian network inference/informant accuracy models for elicitation techniques, the potential applications for robustness testing bear further investigation.

3.3. Concept validity of the criterion graph

> "If a group of 10 persons were all asleep and each person were dreaming of talking to at least one other person in the group, then is there a group structure to be uncovered?" (Killworth and Bernard, 1979).

As we have noted, the modeling approach utilized here depends critically upon the assumption that there exists *some* unique structure which accounts for the commonalities in informant responses. Rather than enmeshing ourselves in a larger debate about what it means to speak of "real" social structures (and whether some might be more "real" than others), we have taken the more restrained position of asserting the notion of the criterion graph purely as a useful construct which is hypothesized to account for shared variance. Even this, however, does not entirely extricate us from the dilemma raised by the Killworth and Bernard comment above. If we find that we are dealing with relations which are purely ascriptive—which are defined purely in terms of actor reports—can we say that we are dealing with social networks at all? From the cognitivist perspective, the answer to this question is clearly yes: if the perceptions (or, from a slightly more behavioristic stance, reports) of individuals are of potential interest to us, then the question of whether an independently verifiable criterion structure can be said to exist is irrelevant. This raises some difficult

[19] Indeed, the most comprehensive reference on network analysis techniques (Wasserman and Faust, 1994) contains no explicit discussion of missing data techniques.

[20] Under the common assumption that no tie is an acceptable "default" condition for dyads.

epistemological (and hence methodological) questions, however. When *is* it sensible to speak of the existence (in at least a hypothetical sense) of the criterion graph? The question is not an idle one. If we, for instance, apply models which presume informant reports to be related to a central structure in situations for which such an assumption is invalid, then the inferences drawn from such an application will he highly misleading at best. On the other hand, ignoring the possibility of a criterion structure where one may be reasonable asserted may substantially limit our ability to draw predictive inferences regarding the social world. As Bernard et al. (1979) note, many processes of interest (e.g. diffusion of information or disease) depend on behavioral, not cognitive, networks. Despite the gloomy prognosis of BKS, work by Romney et al. (1986), Romney and Faust (1982), and Freeman (1992) (among others) suggests that it should be possible to extract at least some useful information about such networks from informant reports. The present work is in this tradition, but recommends caution: we do not consider the existence of the criterion graph to be a trivial assumption, and recognize that the applicability of our approach depends upon the validity of the criterion concept. Further theoretical and empirical development of the foundations of network analysis per se—and, most importantly, of the conceptual foundations of our proposed subject matter—will be necessary if fruitful methodological work in this area is to continue.

4. Conclusion

The dual problems of network inference and informant accuracy pose central methodological challenges for network analysis. While the particular approach employed in dealing with these problems depends on underlying epistemological assumptions, models which assume that informant reports stem from a single criterion graph may be useful in a wide range of circumstances. Given the dimensionality and data efficiency challenges posed by simultaneous estimation of the criterion graph and individual accuracy parameters, a hierarchical Bayesian approach has much to recommend it. Here, we have developed a family of such models, and have shown how they may be applied to the analysis of network data, both for the purposes of direct estimation and for the quantification of uncertainty in derived quantities such as network indices. We have discussed the assumptions implicit in the use of these or other models, and have suggested data collection strategies which will facilitate estimation of network variables. Clearly, the network inference/informant accuracy problem is a serious one, which will require a combination of theoretical, methodological, and empirical research to address. It is hoped that the present work will contribute to this development by providing a suite of theoretically motivated methods with clear applicability to basic problems of empirical research in the field of network analysis.

References

Anderson, B.S., Butts, C.T., Carley, K.M., 1999a. The interaction of size and density with graph-level indices. Social Networks 21 (3), 239–267.

Anderson, C., Wasserman, S., Crouch, B., 1999b. A p^* primer: Logit models for social networks. Social Networks 21, 37–66.

Bailey, K.D., 1987. Methods of Social Research. The Free Press, New York.

Banks, D., Carley, K.M., 1994. Metric inference for social networks. Journal of Classification 11 (1), 121–149.
Batchelder, W.H., Romney, A.K., 1988. Test theory without an answer key. Psychometrika 53 (1), 71–92.
Bernard, H.R., Killworth, P., Sailer, L., 1979. Informant accuracy in social networks. Part IV. A comparison of clique-level structure in behavioral and cognitive network data. Social Networks 2, 191–218.
Bernard, H.R., Killworth, P., 1977. Informant accuracy in social network data. Part II. Human Communication Research 4 (1), 3–18.
Bernard, H.R., Killworth, P., Kronenfeld, D., Sailer, L., 1984. The problem of informant accuracy: the validity of retrospective data. Annual Review of Anthropology 13, 495–517.
Butts, C.T., Carley, K.M., 2001. Multivariate Methods for Interstructural Analysis. CASOS Working Paper. Center for the Computational Analysis of Social and Organization Systems, Carnegie Mellon University.
Calloway, M., Morrissey, J.P., Paulson, R.I., 1993. Accuracy and reliability of self-reported data in interorganizational networks. Social Networks 15, 377–398.
Carley, K.M., Krackhardt, D., 1996. Cognitive inconsistencies and non-symmetric friendship. Social Networks 18, 1–27.
DeVellis, R.F., 1991. Scale Development: Theory and Applications. Applied Social Research Methods Series, vol. 26. Sage Publications, Newbury Park, CA.
Evans, M., Guttman, I., Haitovsky, Y., Swartz, T. B., 1995. Bayesian analysis of binary data subject to misclassification. In: Berry, D., Chaloner, K.M., Geweke, J. (Eds.), Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner. Wiley, New York, pp. 66–77.
Evans, M., Hastings, N., Peacock, B., 2000. Statistical Distributions, 3rd ed. Wiley, New York.
Frank, O., 1988. Random sampling and social networks: a survey of various approaches. Mathematiques, Informatique, et Sciences Humaines 26, 19–33.
Freeman, L.C., 1992. Filling in the blanks: a theory of cognitive categories and the structure of social affiliation. Social Psychology Quarterly 55 (2), 118–127.
Freeman, L.C., Romney, A.K., Freeman, S.C., 1987. Cognitive structure and informant accuracy. American Anthropologist 89, 310–325.
Gamerman, D., 1997. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall, London.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, London.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–511 (with discussion).
Gentle, J.E., 1998. Random Number Generation and Monte Carlo Methods. Springer, New York.
Hammer, M., 1985. Implications of behavioral and cognitive reciprocity in social network data. Social Networks 7, 189–201.
Hildum, D.C., 1986. “Competence” and “performance” in network structure. Social Networks 8, 79–95.
Killworth, P.D., Bernard, H.R., 1976. Informant accuracy in social network data. Human Organization 35 (8), 269–286.
Killworth, P.D., Bernard, H.R., 1979. Informant accuracy in social network data. Part III. A comparison of triadic structure in behavioral and cognitive data. Social Networks 2, 10–46.
Krackhardt, D., 1987a. Cognitive social structures. Social Networks 9, 109–134.
Krackhardt, D., 1987b. QAP partialling as a test of spuriousness. Social Networks 9, 171–186.
Krackhardt, D., 1988. Predicting with networks: non-parametric multiple regression analyses of dyadic data. Social Networks 10, 359–382.
Krackhardt, D., 1990. Assessing the political landscape: structure, cognition, and power in organizations. Administrative Science Quarterly 35, 242–260.
Krackhardt, D., Kilduff, M., 1999. Whether close or far: perceptions of balance in friendship networks in organizations. Journal of Personality and Social Psychology 76 (5), 770–782.
Marsden, P.V., 1990. Network data and measurement. Annual Review of Sociology 16, 435–463.
Pattison, P., Wasserman, S., 1999. Logit models and logistic regressions for social networks. Part II. Multivariate relations. British Journal of Mathematical and Statistical Psychology 52, 169–193.
Pattison, P., Wasserman, S., Robins, G., Kanfer, A.M., 2000. Statistical evaluation of algebraic constraints for social networks. Journal of Mathematical Psychology 44, 536–568.
Robert, C.P., 1994. The Bayesian Choice: A Decision-Theoretic Motivation. Springer, New York.

Robins, G., Pattison, P., Wassernan, S., 1999. Logit models and logistic regressions for social networks. Part III. Valued relations. Psychometrika 64, 371–394.

Romney, A.K., Faust, K., 1982. Predicting the structure of a communications network from recalled data. Social Networks 4, 285–304.

Romney, A.K., Weller, S.C., Batchelder, W.H., 1986. Culture as consensus: a theory of culture and informant accuracy. American Anthropologist 88 (2), 313–338.

Schwarz, N., 1999. Self-reports: how the questions shape the answers. American Psychologist 54 (2), 93–105.

Schwarz, N., Groves, R.M., Schuman, H., 1998. Survey methods. In: Gilbert, D.T., Fiske, S.T., Lindzey, G. (Eds.), The Handbook of Social Psychology, 4th ed., vol. 1. Oxford University Press, New York.

Snijders, T.A.B., 2002. Markov Chain Monte Carlo estimation of exponential random graph models. Journal of Social Structure 3 (2).

Thompson, S.K., Frank, O., 2000. Model-based estimation with link-tracing sampling designs. Survey Methodology 26 (1), 87–98.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.

Wasserman, S., Pattison, P., 1996. Logit models and logistic regressions for social networks. Part I. An introduction to Markov graphs and p^*. Psychometrika 60, 401–426.