Published in final edited form as:

J Comput Graph Stat. 2012 December 1; 21(4): 856–882.

Computational Statistical Methods for Social Network Models

David R. Hunter [Professor],

Department of Statistics, Pennsylvania State University, University Park, PA
(dhunter@stat.psu.edu)

Pavel N. Krivitsky [Research Associate], and

Department of Statistics, Pennsylvania State University, University Park, PA
(krivitsky@stat.psu.edu)

Michael Schweinberger [Research Associate]

Department of Statistics, Pennsylvania State University, University Park, PA
(michael.schweinberger@stat.psu.edu)

Abstract

We review the broad range of recent statistical work in social network models, with emphasis on computational aspects of these methods. Particular focus is applied to exponential-family random graph models (ERGM) and latent variable models for data on complete networks observed at a single time point, though we also briefly review many methods for incompletely observed networks and networks observed at multiple time points. Although we mention far more modeling techniques than we can possibly cover in depth, we provide numerous citations to current literature. We illustrate several of the methods on a small, well-known network dataset, Sampson's monks, providing code where possible so that these analyses may be duplicated.

Keywords

Degeneracy; ERGM; Latent variables; MCMC MLE; Variational methods

1. INTRODUCTION

A typical statistical data frame includes sampling units, which may be considered individuals, and analysis often focuses on some property of these units. Loosely speaking, social networks arise whenever the “property” of interest involves interactions between multiple sampling units, rather than the units themselves. We do not limit ourselves to the case in which the sampling units are actually human beings, though this is by far the most common application that has appeared in the literature on social network models.

There is a long history of work that may be characterized as related to social networks—as Carrington and Scott (2011) pointed out, it is difficult to pinpoint the genesis of this field but

© 2012 American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America

All three authors contributed equally to this article.

Publisher's Disclaimer: Full terms and conditions of use: <http://amstat.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

its roots may be traced at least as far back as the 1930s—though we do not focus on this development here, both because there already exist numerous treatises on networks in general and social networks in particular and because for the audience of *Journal of Computational and Graphical Statistics (JCGS)*, we wish to focus on computational questions. However, we can at least give a partial list of survey-type references for readers interested in delving into the subject of social networks more deeply. Though almost two decades old, the classic book by Wasserman and Faust (1994) is still considered a comprehensive introduction to the important quantitative concepts of social network analysis. As to statistical analysis for social networks, more recent works include the survey article by Goldenberg et al. (2009) and the book-length treatment of various network-related statistical topics by Kolaczyk (2009), both of which give numerous references. Finally, we recommend the other network-related articles in this issue.

This article highlights some current topics in social network modeling, with special emphasis on computational aspects. In Section 2, we introduce a classic dataset, which, though extremely small especially by modern standards, serves to illustrate some of these computational techniques even if it does not demonstrate the state of the art in computational techniques designed for massive social networks. To keep the article to a manageable length, in Section 3, we merely highlight many important topics illustrating the range of statistical work on social network applications, citing recent references to enable interested readers to learn more. The topics that we describe in more detail in Sections 4 and 5 share a common feature: they focus on complete networks that are cross-sectional, which means that they are observed at only one point in time. The distinction between the techniques in Section 4 and those in Section 5 is that the former covers network models whose dependence structure does not have a clear hierarchy, typically expressed as a joint distribution of edge variables, and often focused on modeling global network features and social forces, whereas the latter covers network models that are hierarchical in nature, where the edge variable distributions are parameterized in terms of latent variables, and are focused on identifying individual nodes' roles and positions. We conclude in Section 6 with a discussion of some future challenges. Our hope throughout is to stimulate interest among the readership of *JCGS* to delve into the rapidly expanding field of statistical modeling of social networks.

2. DATA AND NOTATION

We introduce some notation used throughout the article and then discuss a classic social network dataset, which we use as an illustrative example throughout the article.

2.1 Notation

Let N be the set of nodes in the network of interest, indexed $\{1, \dots, n\}$. The relationships in the network may be directed (e.g., friendship nominations, messages) or undirected (e.g., sexual partnerships, conversations). In the former case, we define the set of *dyads* (here used to refer to potential relationships) $\mathbb{Y}$ to be a subset of $N \times N$, the set of ordered pairs of nodes; in the latter case, it is a subset of $\{\{i, j\} : (i, j) \in N \times N\}$, the unordered pairs of nodes. (We will also use $u(\mathbb{Y})$ to refer to an “unordered” version of $\mathbb{Y}$ i.e., $u(\mathbb{Y}) \equiv \{\{i, j\} : (i, j) \in \mathbb{Y}\}$.) Usually, $\mathbb{Y}$ is further constrained in that in most social networks studied, a node cannot have a relationship of interest with itself, excluding pairs of the form (i, i) .

For binary networks, in which the relationship of interest must be either present or absent, we use $\mathcal{Y} \subseteq 2^{\mathbb{Y}}$, the set of subsets of $\mathbb{Y}$ to refer to the set of possible networks of interest, which may be further constrained (i.e., $\mathcal{Y}$ may be a proper subset of $2^{\mathbb{Y}}$). We will use $\mathbf{Y}$ to refer to network random variables and $\mathbf{y} \in \mathcal{Y}$ to refer to their realizations, and $\mathbf{y}_{i,j}$ shall be a 0–1 indicator of whether a relationship of interest is present between i and j in a binary network context.

2.2 Data

The dataset collected by Sampson (1968) and described by Batagelj and Mrvar (2003) is a classic dataset in social network analysis. The dataset summarizes relationships, observed at three distinct time points, among 18 monks who were about to enter a monastery when a conflict erupted. We use here the directed network where $y_{i,j} = 1$ denotes that monk i liked monk j at any of the three time points and $y_{i,j} = 0$ otherwise. The directed network is shown in Figure 1, where circles represent monks and directed edges are oriented from i to j whenever $y_{i,j} = 1$. The monks were divided by Sampson into three groups: Loyal Opposition, Turks, and Outcasts.

3. RANGE OF SOCIAL NETWORK MODELS

The range of statistical modeling techniques for social networks is too broad to address adequately in a single article. We have chosen to focus in some depth in Sections 4 and 5 on cross-sectional (i.e., observed once only), completely observed networks. The current section, by contrast, seeks to illuminate the myriad other applications by describing several additional important recent trends in social network modeling for which we do not have space for a lengthier exposition.

3.1 Dynamic Markovian Models of Networks

The modeling of social network dynamics—that is, changes over time—has attracted much attention, starting with the ground-breaking work of Holland and Leinhardt (1977a,b) and Wasserman (1977, 1979, 1980). Holland and Leinhardt argued that continuous-time Markov processes with state space $\mathcal{Y}$ are natural models of social network dynamics. Wasserman's work, later sharpened by Leenders (1995), studied maximum likelihood estimation for continuous-time Markov process models based on the assumptions that dyad processes are independent and stationary.

This early work was expanded by Snijders (2001), who introduced parameterizations of continuous-time Markov processes that allow dyad processes to be dependent and relaxed the restrictive stationarity assumption, provided two or more discrete-time observations of the process are available. Motivated by the work of McFadden (1974) on random utility models, Snijders' parameterizations have the advantage that the Markov process can be interpreted as “actor-driven,” that is, driven by nodes that maximize random utility functions of the network. More importantly from the standpoint of statistical computing, Snijders (2001) also adapted the method of simulated moments (McFadden 1989) to method of moments estimation of continuous-time Markov models, implemented by stochastic approximation (Robbins and Monro 1951; Pflug 1996). Some computational improvements were discussed by Schweinberger and Snijders (2007), and maximum likelihood and Bayesian estimation were proposed by Snijders, Koskinen, and Schweinberger (2010) and Koskinen and Snijders (2007), respectively. These computational methods are based on nonstandard (Markov chain) Monte Carlo data-augmentation methods and are implemented in the Windows-based program SIENA (Snijders et al. 2012) and the platform-independent R package RSiena (Ripley and Snijders 2011).

More recently, discrete-time Markov models have been explored as alternatives to continuous-time Markov models. Hanneke, Fu, and Xing (2010) explored discrete-time Markov models in which the transition probabilities are expressed by exponential-family random graph models (ERGMs). Krivitsky and Handcock (2012) proposed separable parameterizations of discrete-time Markov models, where one process governs the addition of edges and the other process governs the deletion of edges at each time step; their estimation methods, which extend the (Markov chain) Monte Carlo maximum likelihood methods of Geyer and Thompson (1992), are implemented in the R package ergm (Handcock et al. 2012).

3.2 Dynamic Non-Markovian Models of Networks

Several non-Markovian models for changing networks, including log-linear models and models in which changes are selected uniformly conditional on the degree structure of the network, were discussed briefly by Frank (1991), who cites multiple references. A more recent method is the random-effects model suggested by Westveld and Hoff (2011), which may be adapted to network data that are either binary or in which edges between nodes have normally distributed weights.

One recent trend in the non-Markovian vein in which computation plays an increasingly important role is the application of survival analysis to continuously collected network data (Butts 2008; Brandes, Lerner, and Snijders 2009), an increasingly common paradigm with Internet-based and other computer-generated network datasets. If an “event” is the formation of a new edge, then we attach a counting process either to every node or to every pair of nodes. Vu et al. (2011a) referred to these as the “egocentric” and “relational” models, respectively. The analysis of large datasets (with thousands of nodes and tens of thousands of edges) based on these counting processes is demonstrated in the egocentric case by Vu et al. (2011a) and in the relational case by Vu et al. (2011b) and Perry and Wolfe (2011).

3.3 Joint Models of Networks and Other Outcome Variables

There is a growing literature in which statistical models for the edges in a network are only one part of a larger joint model. Computing plays a huge role in this area, as simulations of these joint models are vital; yet statistical inference is a relatively recent addition to the computational mix. Snijders, Steglich, and Schweinberger (2007), for instance, modeled jointly a dynamically evolving network along with behavior measured on the nodes in that network; parameter estimation for such models given longitudinally observed data is implemented in SIENA (Snijders et al. 2012; Ripley and Snijders 2011). Fellows and Handcock (2012) demonstrated the computational viability of models in which covariates on the nodes in a network are considered random variables. An important subclass of these joint models are models in which a social network is modeled together with some other process of interest that takes place among the nodes in the network, as in the study of infectious diseases. For instance, Britton and O’Neill (2002), whose models were later extended and applied by Groendyke, Welch, and Hunter (2011, 2012), used a Markov chain Monte Carlo (MCMC)-based Bayesian approach to infer model parameters for a contact network underlying an outbreak of an infectious disease when certain information is available about the outbreak, even in cases where none of the contact network edges is actually observed.

3.4 Measurement Error Models and Partially Sampled Networks

Accuracy of network data may be suspect for a number of reasons. A vast sociological literature on the question of how reliable respondents’ reports might be has failed to resolve the question; yet even if it is not quite true that “cognitive data (i.e., recall of who one talks to) ... may not be used for drawing any conclusions about behavioral social structure [i.e., who one actually talks to]” (Bernard, Killworth, and Sailer 1979, p. 191), the statistical community will certainly not be surprised to learn that noise in data collection, along with missing data, can pose significant problems in data analysis. Butts (2003) explored this theme from a statistical point of view, advancing a hierarchical Bayesian framework for simultaneously analyzing network structure along with respondent accuracy. Wyatt, Choudhury, and Bilmes (2008) reported promising results inferring parameters for a network model from noisy data using a stochastic gradient ascent algorithm to optimize the computationally intractable likelihood function.

An important subcategory of models for which data are imperfectly observed are those in which some (known) portion of the network observations is missing. Both maximum likelihood (Gile and Handcock 2010) and Bayesian (Koskinen, Robins, and Pattison 2010) approaches to this

missing-data problem have been proposed and demonstrated to be viable, though each relies heavily on computational techniques. Both use MCMC to help approximate an intractable likelihood, while the latter also uses Bayesian data augmentation.

4. EXPONENTIAL-FAMILY RANDOM GRAPH MODELS: GLOBAL NETWORK CHARACTERISTICS

In this section, we discuss models represented in terms of the joint distribution of all edges, as opposed to a series of conditional distributions of a hierarchical model as described in Section 5. The most popular framework for these joint models is the class of ERGMs. Originally proposed by Holland and Leinhardt (1981) to model individual heterogeneity of nodes and reciprocity of their edges (called the p_1 model), the framework was generalized by Frank and Strauss (1986), Wasserman and Pattison (1996) (who called it the p^* model), and Snijders et al. (2006). It takes the form of a (curved) exponential family on the sample space $\mathcal{Y}$

$$P_{\theta}(Y=y) = \frac{\exp\{\eta(\theta)^{\top} g(y)\}}{\kappa(\theta)}, y \in \mathcal{Y}, \quad (1)$$

where θ is a q -vector of model parameters, which are mapped to a p -vector of natural parameters by $\eta(\cdot)$, and $g(\cdot)$ is a p -vector of sufficient statistics, which capture network features of interest, its postulated dependence structure, or both. We present some examples of ERGM statistics in Table 1. Finally, to make all probabilities sum to one,

$$\kappa(\theta) = \sum_{y' \in \mathcal{Y}} \exp\{\eta(\theta)^{\top} g(y')\}.$$

Notably, this differs from the “textbook” exponential family formula, in that it omits an “ $h(y)$ ” factor, which, together with $\mathcal{Y}$ controls the “reference measure” for the model, the distribution of networks when $\eta(\theta) = \mathbf{0}$. While usually omitted in binary ERGMs, it gains a great deal of importance when using exponential families to model valued networks ties (Krivitsky 2012).

Exponential families for networks can be derived by postulating a substantively relevant conditional dependence structure, and based on this structure, approaches like the Hammersley-Clifford Theorem (Besag 1974) can be used to derive sufficient statistics associated with the model. Frank and Strauss (1986) derived a set of statistics under the “Markovian” assumption that the states of two relationships are, conditional on the rest of the network, stochastically dependent only if they have at least one node in common. Other examples of this approach include the work of Pattison and Wasserman (1999) on multivariate relations; Robins, Pattison, and Wasserman (1999) on polytomous relations; and the realization-dependent conditional independence of Snijders et al. (2006). At the same time, choice of statistics can be driven by beliefs about the factors that influence the underlying social process, without regard for dependence structure (Morris, Handcock, and Hunter 2008; Krivitsky 2012).

For each dyad, one can derive its conditional distribution given the rest of the network, that is,

$$\text{Odds}_{\theta}(Y_{i,j}=1|Y_{-(i,j)}=y_{-(i,j)}) = \frac{\exp\{\eta(\theta)^{\top} g(y_{+(i,j)})\} / \overline{\kappa(\theta)}}{\exp\{\eta(\theta)^{\top} g(y_{-(i,j)})\} / \overline{\kappa(\theta)}} = \exp\{\eta(\theta)^{\top} \Delta_{i,j} g(y)\},$$

with $\Delta_{i,j} \mathbf{g}(\mathbf{y}) \equiv \mathbf{g}(\mathbf{y} \cup \{(i, j)\}) - \mathbf{g}(\mathbf{y} \setminus \{(i, j)\})$, the change in the sufficient statistic vector associated with adding an edge at (i, j) . These “change statistics” (Hunter and Handcock 2006) or “change scores” (Snijders et al. 2006) facilitate “local” interpretation of ERGMs. For example, consider the mutual dyads statistic in Table 1: $\theta_{\leftrightarrow}$ associated with this statistic can be said to increase the conditional odds of an edge (i, j) by $\exp\{\theta_{\leftrightarrow}\}$ if there is already a tie from j to i .

When $\Delta_{i,j} \mathbf{g}(\mathbf{y})$ does not depend on $\mathbf{y}$, the model is dyad independent, and can be decomposed into logistic regression for each edge.

4.1 Simulation Methods

Many applications of these models, including all of the Monte-Carlo-based methods for finding a maximum likelihood estimator (MLE) as well as Bayesian methods, require making draws from the ERGM.

Because evaluating the conditional probability of an edge given the rest of the network is often relatively inexpensive, it is straightforward to simulate network realizations even from intractable ERGMs using a Metropolis–Hastings sampling procedure: given a proposal $\mathbf{y}'$ from density $q(\mathbf{y}' | \mathbf{y})$, accept with probability

$$\frac{q(\mathbf{y}|\mathbf{y}')\exp\{\eta(\theta)^\top \mathbf{g}(\mathbf{y}')\}/\kappa(\theta)}{q(\mathbf{y}'|\mathbf{y})\exp\{\eta(\theta)^\top \mathbf{g}(\mathbf{y})\}/\kappa(\theta)} = \frac{q(\mathbf{y}|\mathbf{y}')}{q(\mathbf{y}'|\mathbf{y})} \exp\{\eta(\theta)^\top (\mathbf{g}(\mathbf{y}') - \mathbf{g}(\mathbf{y}))\}$$

or 1, whichever is smaller. Supposing that $q(\mathbf{y}' | \mathbf{y}) > 0$ only if $\sum_{(i,j) \in \mathcal{Y}} \mathbb{I}(\mathbf{y}'_{i,j} \neq \mathbf{y}_{i,j})$ equals one—that is, the proposed network is constructed by toggling a single edge, say (i, j) —then $\mathbf{g}(\mathbf{y}') - \mathbf{g}(\mathbf{y})$ reduces to $\pm \Delta_{i,j} \mathbf{g}(\mathbf{y})$, with the sign depending on the direction of the toggle, facilitating a fast Gibbs sampling algorithm. Morris, Handcock, and Hunter (2008) further optimized it for large, sparse networks by implementing an asymmetric proposal.

The rate of convergence for this Gibbs procedure was studied by Bhamidi, Bresler, and Sly (2008). There is also some work in progress on exact sampling (Butts 2012).

4.2 Inference Methods

The major challenge associated with applying these models is the intractability of the normalizing constant $\kappa(\theta)$ in the likelihood. Here, we describe the methods used to address this challenge.

4.2.1 Pseudo-Likelihood Estimation—An approximate approach to maximum likelihood estimation is based on the so-called pseudo-likelihood function (Strauss and Ikeda 1990), defined as

$$\prod_{(i,j) \in \mathcal{Y}} P_{\theta}(Y_{i,j} = \mathbf{y}_{i,j} | \mathbf{Y} \setminus \{(i, j)\} = \mathbf{y} \setminus \{(i, j)\}) \prod_{(i,j) \in \mathcal{Y}} \frac{1}{1 + \exp\{-\theta^\top \Delta_{i,j} \mathbf{g}(\mathbf{y})\}} \approx P_{\theta}(\mathbf{Y} = \mathbf{y}).$$

In other words, for each potential edge in the network, its conditional probability given the state of the rest of the network is evaluated, and the product of those probabilities is used to approximate the likelihood. Maximizing the pseudo-likelihood results in a maximum pseudo-likelihood estimator (MPLE), which, computationally, reduces to logistic regression (Strauss and Ikeda 1990). van Duijn, Gile, and Handcock (2009) and others showed that the MPLE is

often biased and far less efficient than the MLE, particularly when the social process modeled has strong dyadic dependence and is particularly vulnerable to the “degeneracy” issues discussed below. van Duijn, Gile, and Handcock (2009) also proposed a maximum bias-corrected pseudolikelihood estimator (MBLE).

4.2.2 Maximum Likelihood Estimation: Stochastic Approximation—In the case of linear ERGMs, the MLE of θ solves

$$\nabla_{\theta} \log P_{\theta}(Y=y) = g(y) - E_{\theta}(g(Y)) = 0, \quad (2)$$

where $\nabla_{\theta} \log P_{\theta}(Y=y)$ denotes the gradient of the log-likelihood function $\log P_{\theta}(Y=y)$ with respect to θ . Snijders (2002) proposed solving Equation (2) using stochastic approximation (Robbins and Monro 1951; Pflug 1996). Starting with an initial guess, the stochastic approximation method updates θ^t to θ^{t+1} at iteration $t+1$ as follows:

$$\theta^{t+1} = \theta^t - a_t \hat{D}_t(g(Y_{\theta^t}) - g(y)),$$

where D_t^{-1} is an approximation of the gradient $\nabla_{\theta} E_{\theta}(g(Y))$ in the neighborhood of θ^t , a_t is a sequence of numbers which tends sufficiently slowly to 0 as t increases, and Y_{θ^t} is a network sample from the ERGM with parameter θ^t by MCMC methods. Okabayashi and Geyer (2012) proposed a linear search algorithm along similar lines, while Jin and Liang (2012) developed yet another stochastic approximation algorithm.

4.2.3 Maximum Likelihood Estimation: Monte Carlo Maximization—An alternative approach to maximum likelihood estimation is based on Monte Carlo approximations of the likelihood function (1). The MCMC MLE approach of Geyer and Thompson (1992) was first adapted by Handcock (2003) to ERGMs and extended by Hunter and Handcock (2006). Both the stochastic approximation algorithm of Snijders (2002) described in Section 4.2.2 and the MCMC MLE approach of Hunter and Handcock (2006) rely on MCMC simulations of networks, but the MCMC MLE approach has the advantage that it makes more efficient use of MCMC samples, as pointed out by Geyer and Thompson (1992, sec. 1.3).

Starting with an initial guess θ , which is often taken to be the easy-to-calculate MPLE, the ratio of intractable normalizing constants is approximated as

$$\frac{\kappa(\theta')}{\kappa(\theta)} = \sum_{y \in \mathcal{Y}} \exp\{(\eta(\theta') - \eta(\theta))^T g(y)\} \frac{\exp\{\eta(\theta)^T g(y)\}}{\kappa(\theta)} = E_{\theta}(\exp\{(\eta(\theta') - \eta(\theta))^T g(Y)\}),$$

with the last expectation approximated by a sample of sufficient statistics under θ , allowing the likelihood to be maximized as a function of θ' .

In practice, the accuracy of the approximation decreases as θ' moves farther from θ ; in particular, if a guess θ is so far from the MLE that the interior of the convex hull of a sample under θ does not contain the observed sufficient statistic, the maximized θ' does not exist (Hummel, Hunter, and Handcock 2012). Thus, practical implementation such as that of Handcock et al. (2012) involves several iterations of refining the guess and sampling from it. The MPLE is often used for the purpose, and Okabayashi and Geyer (2012) suggest using their linear search algorithm for the purpose, while Hummel, Hunter, and Handcock (2012) propose an adaptive method to attenuate the change in θ' .

Publicly available implementations of ERGM MLE and MPLE include the `ergm` package (Hunter et al. 2008; Handcock et al. 2012) from the Statnet suite of R packages and `PNet` (Wang, Robins, and Pattison 2009).

4.2.4 Bayesian Methods—Bayesian inference is based on the posterior distribution of θ given y :

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta),$$

where $p(y) = \int p(y | \theta) p(\theta) d\theta$ is the marginal probability of y , $p(y | \theta) = P_{\theta}(Y = y)$ is the conditional probability of y given θ , $p(\theta)$ is the prior probability density of θ , and $p(\theta | y)$ is the posterior density of θ given y . The posterior density $p(\theta | y)$ is typically intractable, because its normalizing constant $p(y)$ is intractable.

Standard MCMC methods, for example, the Metropolis–Hastings algorithm, can deal with intractable normalizing constants of a posterior density as long as the posterior density in question is known up to a constant. The problem is that the posterior density is not known up to a constant, because the normalizing constant $\kappa(\theta)$ of the likelihood function $p(y | \theta)$ is intractable. Since the posterior density $p(\theta | y)$ of complex ERGMs includes two intractable normalizing constants, $p(y)$ on the one hand and $\kappa(\theta)$ on the other hand, the posterior density is *doubly intractable* (Murray, Ghahramani, and MacKay 2006).

To demonstrate that standard MCMC methods cannot sample from doubly intractable posterior densities, suppose that we want to construct a Markov chain with stationary distribution $p(\theta | y)$ by a Metropolis–Hastings algorithm. If θ denotes the current value of the parameter and $\hat{\theta}$ denotes a proposal of the parameter generated from a proposal density $q(\cdot | \theta)$, then $\hat{\theta}$ is accepted with probability $\min(1, a)$, where

$$a = \frac{p(y|\hat{\theta})p(\hat{\theta})}{p(y|\theta)p(\theta)}. \quad (3)$$

The problem is that acceptance probability (3) is intractable, because $p(y | \theta)$ and $p(y | \hat{\theta})$ involve the normalizing constants $\kappa(\theta)$ and $\kappa(\hat{\theta})$, respectively.

A naive approach would be to replace the intractable likelihood ratio by an approximation, either deterministic or stochastic. A deterministic approximation could be based on variational approximations of the log-normalizing constants, while stochastic approximations could be based on MCMC estimators of the ratio of normalizing constants such as the importance sampling estimator used in maximum likelihood estimation (e.g., Murray 2007). However, the stationary distribution of the resulting Markov chains may not be the desired target distribution, that is, the posterior density of interest (Murray 2007).

The problematic nature of the naive approach has led to the development of a body of MCMC methods that by design generate samples from doubly intractable posterior densities. Most of them are based on augmenting the posterior density so that the augmented posterior probability distribution is easy to sample from. We discuss here one of the simplest and most appealing auxiliary-variable MCMC methods, following Murray, Ghahramani, and MacKay (2006) and Caimo and Friel (2011). Other auxiliary-variable approaches are discussed by Møller et al. (2006) and Koskinen, Robins, and Pattison (2010); an alternative approach, inspired by the Monte Carlo maximum likelihood algorithm of Geyer and Thompson (1992), was proposed by Atchadé, Lartillot, and Robert (2012).

The basic idea can be described as follows. The data $\mathbf{y}$ are augmented by an auxiliary random graph $\tilde{\mathbf{Y}}$ and an auxiliary parameter vector $\tilde{\boldsymbol{\theta}}$. Suppose the joint density of $\boldsymbol{\theta}$, $\mathbf{Y}$, $\tilde{\boldsymbol{\theta}}$, $\tilde{\mathbf{Y}}$ is of the form

$$p(\boldsymbol{\theta}, \mathbf{y}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{y}}) = p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{y})p(\tilde{\mathbf{y}}|\tilde{\boldsymbol{\theta}}), \quad (4)$$

where $q(\tilde{\boldsymbol{\theta}} | \boldsymbol{\theta}, \mathbf{y})$ is an auxiliary density and $p(\tilde{\mathbf{y}} | \tilde{\boldsymbol{\theta}})$ is the conditional probability of $\tilde{\mathbf{y}}$ given $\tilde{\boldsymbol{\theta}}$, which is of the same exponential-family form as $\mathbf{Y}$, implying the same reference measure and the same sufficient statistics. The augmented posterior density is of the form

$$p(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{y}}|\mathbf{y}) \propto p(\boldsymbol{\theta}, \mathbf{y}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{y}}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{y})p(\tilde{\mathbf{y}}|\tilde{\boldsymbol{\theta}}). \quad (5)$$

The posterior density of interest, $p(\boldsymbol{\theta} | \mathbf{y})$, is the marginal distribution of the augmented posterior density, $p(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{y}} | \mathbf{y})$. A simple Metropolis–Hastings algorithm to sample from the augmented posterior density, which has been called the *exchange algorithm* (Murray, Ghahramani, and MacKay 2006; Caimo and Friel 2011), operates as follows:

1. Sample $\tilde{\boldsymbol{\theta}} | \boldsymbol{\theta}, \mathbf{y} \sim q(\cdot | \boldsymbol{\theta}, \mathbf{y})$ and then sample $\tilde{\mathbf{Y}} | \tilde{\boldsymbol{\theta}} \sim p(\cdot | \tilde{\boldsymbol{\theta}})$.
2. Propose to swap the values of $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ and accept the proposal with probability $\min(1, a)$, where

$$a = \frac{p(\tilde{\boldsymbol{\theta}})p(\mathbf{y}|\tilde{\boldsymbol{\theta}})q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}, \mathbf{y})p(\tilde{\mathbf{y}}|\boldsymbol{\theta})}{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}, \mathbf{y})p(\tilde{\mathbf{y}}|\tilde{\boldsymbol{\theta}})}. \quad (6)$$

Simple calculation shows that the intractable normalizing constants $\kappa(\boldsymbol{\theta})$ and $\kappa(\tilde{\boldsymbol{\theta}})$ cancel from the acceptance probability (6), so the Metropolis–Hastings algorithm operating on the augmented state space is tractable. Sampling $\tilde{\mathbf{Y}}$ requires exact sampling, which is typically infeasible. Caimo and Friel (2011) proposed to sample $\tilde{\mathbf{Y}}$ instead by MCMC methods, as described in Section 4.1. The work of Liang (2010) on double Metropolis–Hastings algorithms provides some justification for doing so. Last, to reduce model degeneracy of conventional ERGMs, Schweinberger and Handcock (2011) proposed hierarchical ERGMs with local dependence and extended the exchange algorithm to hierarchical ERGMs.

4.3 ERGM Degeneracy

In addition to the intractable normalizing constant, a major difficulty in modeling complex social processes using ERGMs is a phenomenon referred to as “degeneracy,” which arises, for example, when attempting to model processes exhibiting individual heterogeneity in activity level using 2-star statistics and triad closure using triangle counts as suggested in Frank and Strauss (1986). In a contemporaneous publication, Strauss (1986) noted that the Markov graph models Frank and Strauss (1986) had proposed tend to, as the network size increases, concentrate an increasingly large probability mass on an increasingly small fraction of possible networks. Jonasson (1999) and Häggström and Jonasson (1999) studied the asymptotics of the so-called triangle model, corresponding to the ERGM with the number of edges and triangles as sufficient statistics, and showed that as the network size increases, the parameter space tends to be divided between configurations that exhibit little triadic closure—typically less than observed—and configurations that are degenerate, in that almost all of the probability mass is concentrated on a single (complete) network. Handcock (2003) and Rinaldo, Fienberg, and Zhou (2009), among others, mapped the shape of the region of the parameter space in which this degeneracy occurred, and Schweinberger (2011) showed that some statistics, including triangle and k -star counts, induce this behavior asymptotically.

It can be argued that this degeneracy is merely a symptom of a broader problem of model misspecification. In their development of curved ERGMs, Snijders et al. (2006) used change statistics to show that models with positive coefficients on triangles and/or k -stars induced strong positive dependence among dyad values: edges beget more edges, leading to what the authors called an “avalanche” toward a complete graph. In another realization, the avalanche might not take place, leading to a sparse network. Thus, the model with such positive dependence can induce a bimodal distribution of networks, and even at the MLE, it is found that the observed network is between the modes. A bimodal distribution, with little mass between the modes, is difficult for MCMC sampling to explore, which can cause any simulation-based method to fail, whether maximum likelihood or Bayesian.

Indeed, Krivitsky (2012), in developing ERGM statistics for networks with weighted edges that are unbounded counts, reports that a statistic analogous to the count of 2-stars induced a bimodal distribution of networks, with neither mode being degenerate (i.e., concentrated) itself. Thus, a fruitful approach to degeneracy may be to model network phenomena using statistics less prone to result in model degeneracy, such as those of Snijders et al. (2006) and Krivitsky (2012).

4.4 Application

We demonstrate the three methods of estimation by applying them to Sampson’s network dataset and the ERGM with the three statistics of Table 1 except that the transitive triad statistic is replaced by the less degeneracy-prone transitive edges statistic given by

$$\sum_{(i,j) \in \mathbb{Y}} y_{i,j} \max_{k \in N \setminus \{i,j\}} y_{i,k} y_{k,j}$$

(Snijders, van de Bunt, and Steglich 2010). This statistic resembles the transitive triad count in Table 1 except the summation over k is replaced by a maximum over k .

Table 2 shows various point estimates of the edges, mutual edges, and transitive edges parameters. Although there is some question as to the validity of ERGM standard errors due to questionable asymptotics (Hunter and Handcock 2006, e.g.), the standard errors in this case seem to agree closely with the posterior standard deviations, suggesting that they may be valid in practice. The marginal posterior densities are shown in Figure 2, along with the MCMC MLE. It is evident that the MCMC MLE is close to the posterior mode. Example code to obtain the results shown here using the R packages *ergm* and *Bergm* can be found in the Appendix (in both R packages, the transitive edges statistic is called *transitivities*).

Hunter, Goodreau, and Handcock (2008) argued that assessing the goodness of fit of models is important, not least due to the model degeneracy problem discussed in Section 6. We show posterior predictions of the sufficient statistics in Figure 3. The posterior predictions are centered at the observed data, which is encouraging.

5. MODELS BASED ON INDEPENDENCE CONDITIONAL ON LATENT VARIABLES

While ERGMs are useful for modeling global network characteristics, models based on conditional independence (given latent variables) are useful for multiple reasons. First, ERGMs are not well understood and sometimes possess undesirable properties, for example, model degeneracy. Second, the likelihood function of ERGMs can be intractable, complicating statistical computing. Third, there may be unobserved heterogeneity or unobserved structure.

For these reasons, models building on conditional independence (given latent variables) have attracted much attention. At least three streams of latent variables can be distinguished: random effects models and mixed effects models and extensions; stochastic block models and extensions, including mixed membership models; and latent space models and extensions. All of these latent variable models, with the exception of Wyatt, Choudhury, and Bilmes (2008); Koskinen (2009); and Schweinberger and Handcock (2011), assume conditional independence of dyads: if $\mathbf{Z}$ denotes generic latent variables, which may be node-bound or dyad-bound and discrete or continuous, then the models assume that dyads (distinct unordered pairs of nodes) $(Y_{i,j}, Y_{j,i})$ are independent conditional on $\mathbf{Z}$:

$$P_{\theta}(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z}) = \prod_{(i,j) \in \text{eu}(\mathbb{Y})} P_{\theta}(Y_{i,j}=y_{i,j}, Y_{j,i}=y_{j,i}|\mathbf{Z}=\mathbf{z}).$$

In the case of directed networks, conditional independence of dyads $(Y_{i,j}, Y_{j,i})$ does not imply conditional independence of edges $Y_{i,j}$ and $Y_{j,i}$: even conditional on $\mathbf{Z}$, edges $Y_{i,j}$ and $Y_{j,i}$ may be dependent due to reciprocity. However, with the notable exception of the p_2 model (van Duijn 1995; van Duijn, Snijders, and Zijlstra 2004), most latent variable models make the more restrictive assumption that edges $Y_{i,j}$ are independent conditional on $\mathbf{Z}$:

$$P_{\theta}(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z}) = \prod_{(i,j) \in \mathbb{Y}} P_{\theta}(Y_{i,j}=y_{i,j}|\mathbf{Z}=\mathbf{z}). \quad (7)$$

Three aspects of these models are worth noting. First, conditional independence does not imply that latent variable models cannot capture network dependencies of interest. Indeed, some of the more advanced models (notably latent space models) make clever use of latent variables to capture such dependence structure, including mutuality and transitivity. Second, the conditional independence of edges implies that model degeneracy is not an issue, which facilitates the construction of models. Third, the conditional independence of edges has computational advantages in that standard MCMC methods can be used.

We discuss random effects and mixed effects models very briefly in Section 5.1 and then describe two classic latent variable models in more detail: stochastic block models in Section 5.2 and latent space models in Section 5.3.

5.1 Random Effects and Mixed Effects Models

The earliest latent variable model, the random effects model of van Duijn (1995), was motivated by the lack of model parsimony of the so-called p_1 model of Holland and Leinhardt (1981)—which is an ERGM of the form (1) with the in-degrees, out-degrees, and mutual edges as sufficient statistics—and the fact that the p_1 model ignores covariates. Since it can be considered a random effects version of p_1 , the model of van Duijn (1995) is known as the p_2 model. To estimate its parameters, van Duijn (1995) and van Duijn, Snijders, and Zijlstra (2004) exploited the fact that the p_1 model can be represented as a generalized linear model and the p_2 model as a generalized linear mixed model and developed an iterative generalized least squares procedure. Zijlstra, van Duijn, and Snijders (2009) developed Bayesian MCMC methods for the p_2 model.

Hoff (2003, 2005, 2009) introduced multiple generalizations of generalized linear mixed models, along with Bayesian MCMC methods. Some of these computational methods are implemented in the R package *eigenmodel* (Hoff 2012). Krivitsky et al. (2009) extended the latent cluster models of Section 5.3 to include random effects.

5.2 Stochastic Block Models

Stochastic block models of networks, which are related to finite mixture models, were explored by Snijders and Nowicki (1997) and extended by Nowicki and Snijders (2001). These models partition the set of nodes into subsets, called blocks, where conditional on block memberships, edges are independent and the probability of an edge between two nodes depends on the blocks to which the nodes belong. Tallberg (2005) described an extension that incorporates covariates to predict block memberships. Airoldi et al. (2008) proposed more advanced stochastic block models, called mixed membership models, that allow the block memberships of nodes to depend on the pair of nodes, that is, the block to which one belongs depends on with whom one interacts.

Yet we focus here on the basic stochastic block model of Nowicki and Snijders (2001), which is based on two fundamental assumptions. First, the set of nodes is partitioned into K blocks, where K is fixed and known. Let

$$\mathbf{Z}_i | \pi_1, \dots, \pi_K \stackrel{\text{iid}}{\sim} \text{Multinomial}(1; \pi_1, \dots, \pi_K). \quad (8)$$

Second, Equation (7) holds, with $P_{\boldsymbol{\theta}}(Y_{i,j} = 1 | \mathbf{Z}_i = \mathbf{z}_i, \mathbf{Z}_j = \mathbf{z}_j) = \boldsymbol{\theta}_{\mathbf{z}_i, \mathbf{z}_j}$, where $\boldsymbol{\theta}_{k,l}$ is the probability that a given actor in group k has a tie to a given actor in group l .

The Bayesian MCMC algorithm is simple: given conjugate priors, the posterior can be sampled by Gibbs sampling. In particular, if the priors for $\boldsymbol{\pi}$ and $\boldsymbol{\beta}$ are independent and given by Dirichlet and beta, respectively, then the full conditional distributions of $\boldsymbol{\pi}$ and $\boldsymbol{\beta}$ are also Dirichlet and beta. The parameters of stochastic block models are not identifiable in that the likelihood function is invariant to the labeling of the blocks and Bayesian MCMC samples may therefore show evidence of label switching. Snijders and Nowicki (1997) and Nowicki and Snijders (2001) dealt with the label-switching problem either by imposing order restrictions on parameters or by restricting attention to functions that are invariant to the labeling of the blocks, for example, an indicator of whether two nodes belong to the same block. An alternative is given by the Bayesian decision-theoretic approach of Stephens (2000), which relabels the Bayesian MCMC sample. The first approach is implemented in the Windows-based program BLOCKS (Snijders and Nowicki 2007) and the second approach is implemented in the R package `hergm` (Schweinberger 2012a).

The variational methods of Section 5.4 provide an alternative to MCMC-based estimation. They are approximate but fast and feasible, and they can be applied to networks with more than 100,000 nodes and 10 billion dyads.

5.3 Latent Space Models

Hoff, Raftery, and Handcock (2002) proposed latent space models based on the assumption that nodes are embedded in a latent, metric space. Latent space models come in two basic flavors: one that assumes the metric space is Euclidean (Hoff, Raftery, and Handcock 2002), and another that assumes the metric space is ultrametric (Schweinberger and Snijders 2003). In either case, Equation (7) is assumed to hold with

$$\text{logit}(P_{\boldsymbol{\beta}}(Y_{i,j} = 1 | \mathbf{Z}_i = \mathbf{z}_i, \mathbf{Z}_j = \mathbf{z}_j)) = \boldsymbol{\beta}_0 + \mathbf{x}_{i,j}^{\top} \boldsymbol{\beta} + d(\mathbf{z}_i, \mathbf{z}_j),$$

where $\mathbf{x}_{i,j}$ is a q -vector of covariates for dyad (i, j) , $\boldsymbol{\beta}_0$ is a parameter that controls the density of the network, and $d(\cdot, \cdot)$ is a distance function. Two examples of distance functions are given by $d(\mathbf{z}_i, \mathbf{z}_j) = -\|\mathbf{z}_i - \mathbf{z}_j\|$ for a Euclidean latent effect and $d(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{z}_i^{\top} \mathbf{z}_j$ for an inner-product effect.

The likelihood function, which is equivalent to the likelihood function of nonlinear logistic regression models, can be maximized to obtain maximum likelihood estimates. An alternative is a Bayesian approach, specifying a prior distribution for β and Z and using the Gibbs sampler to sample from the posterior distribution of β and Z . If the posterior distribution is approximated by Bayesian MCMC methods, then a nonidentifiability problem arises, in that the likelihood is invariant to rotation and reflection of latent positions about the origin and, for Euclidean latent space models, it is also invariant to translation. Hoff, Raftery, and Handcock (2002) addressed this problem by using a Procrustes transformation to rotate and reflect these posterior draws to be as close as possible to a reference configuration. Shortreed, Handcock, and Hoff (2006) further addressed this problem via a Minimum Kullback–Leibler divergence (MKL) estimate to summarize the posterior positions, finding a configuration $(\beta_{\text{MKL}}, Z_{\text{MKL}})$ that induces a distribution of dyad values closest to their posterior predictive distribution. That is, given two parameter configurations $\theta = (\beta, z)$ and $\theta' = (\beta', z')$ and letting Y_θ be a random network variable induced by θ , so that $P(Y_\theta = y) \equiv P_\theta(Y = y)$, the divergence is given by

$$\text{KL}(Y_{\theta'}, Y_\theta) = E_{Y'|Y_\theta} \left(\log \frac{P_{\theta'}(Y=Y')}{P_\theta(Y=Y')} \right),$$

with the estimate minimizing its expectation under the posterior $\theta \mid Y_{\text{obs}} = y_{\text{obs}}$:

$$\theta_{\text{MKL}} = \arg \min_{\theta} E_{\theta' \mid Y_{\text{obs}} = y_{\text{obs}}} (\text{KL}(Y_{\theta'}, Y_\theta)).$$

Shortreed, Handcock, and Hoff (2006) showed that it is possible to approximate MKL estimates by MCMC.

This model is further extended by Hoff (2005) to add node-specific random sender (δ) and receiver (γ) effects, that is,

$$\text{logit}(P_\beta(Y_{i,j}=1 \mid Z=z_i, Z=z_j)) = \beta_0 + x_{i,j}^\top \beta + d(z_i, z_j) + \delta_i + \gamma_j,$$

with δ and γ having a normal distribution, and by Handcock, Raftery, and Tantrum (2007) to model the latent positions as a mixture of K spherical Gaussian clusters, that is,

$$Z_i \stackrel{\text{iid}}{\sim} \sum_{k=1}^K \pi_k \text{MVN}(\mu, \sigma_Z^2 I),$$

allowing a soft clustering of the nodes in the network to be produced. This leads to another nonidentifiability problem, in that the prior for the latent positions is invariant to permutation of cluster labels, which is resolved using the relabeling algorithm of Stephens (2000).

Alternative computational approaches to approximating the likelihood function include the case-control sampling idea of Raftery et al. (2012) and the variational approach of Salter-Townshend and Murphy (2013). The latter is described in Section 5.4.

5.4 Variational Methods

Bayesian MCMC methods, which represent the gold standard and dominate the practice of latent variable modeling, tend to be slow and cannot be applied to large networks with hundreds of nodes. Variational methods have emerged as a fast and feasible though approximate

alternative to Bayesian MCMC methods. Approximate maximum likelihood estimation of stochastic block models based on variational methods was introduced by Daudin, Picard, and Robin (2008), who presented a variational expectation-maximization (EM) algorithm and applied it to more than 600 nodes. Online versions of Daudin, Picard, and Robin (2008) algorithms were considered by Zanghi et al. (2010) and applied to more than 12,000 nodes. Airoldi et al. (2008) applied a variational EM algorithm to mixed membership models. A generalized variational EM algorithm, which exploits an majorization-minimization (MM) algorithm (Hunter and Lange 2004) to implement the E-step, was proposed by Vu, Hunter, and Schweinberger (2012) and applied to more than 131,000 nodes. The consistency of approximate maximum likelihood estimators was established by Celisse, Daudin, and Pierre (2011), who considered the stochastic block models of Nowicki and Snijders (2001). Approximate Bayesian estimation of stochastic block models can be found by using methods in Schweinberger, Petrescu-Prahova, and Vu (2011) and Vu, Hunter, and Schweinberger (2012). Approximate Bayesian estimation of latent space models was discussed by Salter-Townshend and Murphy (2013) and applied to more than 80 nodes.

We focus here on approximate Bayesian estimation of latent variable models based on variational methods. To keep the discussion manageable, we let $\mathbf{u}$ represent the vector of unknowns—which includes all latent variables and all parameters.

The key obstacle to Bayesian model estimation and model selection is the marginal likelihood $p(\mathbf{y})$, which can be expressed as

$$p(\mathbf{Y}) = \int \frac{p(\mathbf{y}|\mathbf{u})p(\mathbf{u})}{q(\mathbf{u}|\boldsymbol{\theta})} q(\mathbf{u}|\boldsymbol{\theta}) d\mathbf{u}, \quad (9)$$

where $q(\mathbf{u}|\boldsymbol{\theta})$ is a distribution with support $\mathcal{U}$ the space of $\mathbf{u}$, parameterized by a vector of auxiliary parameters $\boldsymbol{\theta}$. The idea of variational methods is to translate the intractable integration problem (9) into a tractable optimization problem by bounding $\log p(\mathbf{y})$ below using a bound that is as tight as possible. To this end, we maximize the right-hand side of

$$\log p(\mathbf{y}) \geq \int \left[\log \frac{p(\mathbf{y}|\mathbf{u})p(\mathbf{u})}{q(\mathbf{u}|\boldsymbol{\theta})} \right] q(\mathbf{u}|\boldsymbol{\theta}) d\mathbf{u}. \quad (10)$$

with respect to $\boldsymbol{\theta}$. It can be shown that the difference between the left- and right-hand side of (10) is equal to the Kullback–Leibler divergence from the auxiliary distribution $q(\mathbf{u}|\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$ to the posterior distribution $p(\mathbf{u}|\mathbf{y})$:

$$\text{KL}(q(\mathbf{u}|\boldsymbol{\theta}); p(\mathbf{u}|\mathbf{y})) = \int \left[\log \frac{q(\mathbf{u}|\boldsymbol{\theta})}{p(\mathbf{u}|\mathbf{y})} \right] q(\mathbf{u}|\boldsymbol{\theta}) d\mathbf{u}. \quad (11)$$

Thus, maximizing the lower bound (10) is equivalent to minimizing the Kullback–Leibler divergence (11).

To make the optimization problem tractable, it is common to choose fully factorized auxiliary distributions of the form

$$q(\mathbf{u}|\boldsymbol{\theta}) = \prod_i q(u_i|\boldsymbol{\theta}_i), \quad (12)$$

where u_i denotes the i th element of the vector of unknowns $\mathbf{u}$ and $q(u_i|\boldsymbol{\theta}_i)$ denotes the marginal auxiliary distribution of u_i , parameterized by a vector of auxiliary parameters $\boldsymbol{\theta}_i$.

A variational EM algorithm to maximize the lower bound can be sketched as follows. Let $LB(\boldsymbol{\vartheta}_Z^{(t)}, \boldsymbol{\vartheta}_\theta^{(t)})$ be the lower bound (10), where $\boldsymbol{\vartheta}_Z$ denotes the vector of auxiliary parameters corresponding to the latent variables, with value $\boldsymbol{\vartheta}_Z^{(t)}$ at iteration t , and $\boldsymbol{\vartheta}_\theta$ denotes the vector of auxiliary parameters corresponding to the parameters $\boldsymbol{\theta}$, with value $\boldsymbol{\vartheta}_\theta^{(t)}$ at iteration t .

E-STEP: Let $\boldsymbol{\vartheta}_Z^{(t+1)}$ be the maximizer of $LB(\boldsymbol{\vartheta}_Z, \boldsymbol{\vartheta}_\theta^{(t)})$ with respect to $\boldsymbol{\vartheta}_Z$.

M-STEP: Let $\boldsymbol{\vartheta}_\theta^{(t+1)}$ be the maximizer of $LB(\boldsymbol{\vartheta}_Z^{(t+1)}, \boldsymbol{\vartheta}_\theta)$ with respect to $\boldsymbol{\vartheta}_\theta$.

By construction, iteration $t + 1$ increases the lower bound $LB(\boldsymbol{\vartheta}_Z, \boldsymbol{\vartheta}_\theta)$:

$$LB(\boldsymbol{\vartheta}_Z^{(t)}, \boldsymbol{\vartheta}_\theta^{(t)}) \leq LB(\boldsymbol{\vartheta}_Z^{(t+1)}, \boldsymbol{\vartheta}_\theta^{(t)}) \quad (13)$$

$$\leq LB(\boldsymbol{\vartheta}_Z^{(t+1)}, \boldsymbol{\vartheta}_\theta^{(t+1)}) \quad (14)$$

Implementation details are discussed in the cited literature.

A variational approach to Bayesian inference approximates the marginal likelihood $p(\mathbf{y})$ and the posterior distribution $p(\mathbf{u} | \mathbf{y})$ and thus tackles both Bayesian model estimation and model selection.

In practice, the most important problem with variational methods is that it is impossible to assess how tight the lower bound (10) is. In the case of fully factorized auxiliary distributions of the form (12), the Kullback–Leibler divergence (11), and thus the tightness of the lower bound (10), is determined by the posterior dependence of the unknowns. In the absence of posterior dependence, equality can be achieved in (10), though in general this is not possible. However, as long as the posterior dependence is weak, the variational approximation of the posterior distribution and the log marginal likelihood may be useful, at least as a starting point of Bayesian MCMC algorithms and when Bayesian MCMC algorithms are too slow.

5.5 Application

To illustrate latent variable models, we apply the stochastic block model of Nowicki and Snijders (2001) and the latent space models of Handcock, Raftery, and Tantrum (2007) and Krivitsky et al. (2009) to Sampson's monk data.

We obtained a Bayesian posterior MCMC sample from the Nowicki and Snijders (2001) model using the Windows-based program BLOCKS (Snijders and Nowicki 2007). Because BLOCKS returns summaries of the posterior distribution rather than a sample from it, we modified the Delphi source code of BLOCKS to obtain an MCMC sample. We handled the label switching in the posterior sample using a stochastic label-switching algorithm described by Schweinberger and Handcock (2011) and implemented in the R package *hergm* (Schweinberger 2012a). MCMC samples from the posterior distributions of the latent space models of Handcock, Raftery, and Tantrum (2007) and Krivitsky et al. (2009) were obtained by the R package *latentnet* (Krivitsky and Handcock 2008), for which the code can be found in the Appendix.

Figure 4 shows the clustering of Sampson's monks by the three models, where we use both two and three dimensions for the Krivitsky et al. (2009) model. It is evident that the clustering is almost the same with all three approaches and agrees with the ground truth—the three-group partition of the monks into Loyal Opposition, Turks, and Outcasts. In fact, there is hardly any uncertainty about the clustering.

6. DISCUSSION

As we have tried to show, applications that combine computational statistics and social network modeling are already very broad, but research on this combination is still quite young relative to these two separate fields. As such, numerous challenges remain.

First, there are myriad open questions regarding theoretical statistics in network modeling. Research in this area is too broad to summarize adequately in a single paragraph, though we may highlight some computational issues related to this theory. There are several recent articles that extend exponential family theory to network models, with particular emphasis on cases in which MLEs may not exist. Geyer (2009) showed how exact linear programming algorithms, implemented in the R package `rcdd` (Geyer 2012), may be used to analyze such cases. There is very little asymptotic theory about the use of ERGMs—indeed, Shalizi and Rinaldo (2011) pointed out that MLEs for certain such models may be shown not even to be consistent as the number of nodes grows without bound—which means that computational techniques for finite-sample statistical inference and for Bayesian inference become important. Monte-Carlo-based testing (Krivitsky 2012, e.g.) is an obvious possibility here, though there exists little work to date on practical implementation of valid inferential techniques in general.

Beyond the theoretical questions, however, perhaps the most important computational issue looming for statistical analysis of social network data is the high-dimensional nature of more and more network datasets. Since the number of edge variables in a network of n nodes scales like $O(n^2)$, mere storage of large networks becomes problematic. Often, social networks are sparse in the sense that many more edge variables are zero (indicating absence of an edge) than one, which means that computational techniques for sparse matrices become relevant. Though software implementing such techniques already exists for many common matrix operations, novel statistical analyses might require novel matrix operations and therefore new algorithms. Furthermore, it is increasingly the case that n itself is quite large—when data are collected automatically by electronic devices from a large population, for example—so algorithms that scale no better than $O(n^2)$ are problematic.

We feel that the interplay of statistical theory and computing in social network modeling applications will also prove to be fertile ground for researchers in statistics to make advances. As a single example, consider testing a null hypothesis against an alternative hypothesis. Standard tests based on the Wald or likelihood ratio test statistics require maximizing the likelihood function under the alternative hypothesis, which may be infeasible or, at the very least, may be more time consuming than maximizing the likelihood function under the null hypothesis. Thus, score tests, which only require maximizing the likelihood function under the null hypothesis, may present an attractive alternative provided clever methods are used to evaluate the score function under the null hypothesis. In the case of dynamic network models of the form Snijders (2001) and Snijders, Steglich, and Schweinberger (2007), frequentist score-type tests (Schweinberger 2012b) have become popular and are widely used. For exponentially parameterized stochastic block models, Bayesian score-type tests are explored by Schweinberger, Petrescu-Prahova, and Vu (2011).

With regard to degeneracy, as discussed in Section 4.3, dependence creates multimodality for degenerate models, which in turn creates issues for computational algorithms that depend on MCMC. Solving these computational issues will not make degeneracy go away; it can help to cure the symptom, but fails to address the root of the problem. In any case, goodness-of-fit measures are indispensable to diagnose degeneracy where it occurs (Hunter, Goodreau, and Handcock 2008).

Acknowledgments

The authors gratefully acknowledge the support by the Office of Naval Research under the MURI program, Award Number N00014-08-1-1015, and the National Institutes of Health, Award Numbers R01 HD068395 and R01 GM083603.

APPENDIX: EXAMPLE CODE

```

ERGM MCMC MLE
library(ergm)
data(sampson)
monks.ergm.mle <- ergm(samplike~edges+mutual+transitivities)
summary(monks.ergm.mle)
ERGM MPLE
library(ergm)
data(sampson)
monks.ergm.mple <- ergm(samplike~edges+mutual+transitivities,
estimate="MPLE")
summary(monks.ergm.mple)
Two-dimensional Euclidean latent space model with three clusters and random
receiver effects
library(latentnet)
data(sampson)
monks.d2G3r <- ergmm(samplike~euclidean(d=2,G=3)+rreceiver)
Z <- plot(monks.d2G3r, rand.eff="receiver," pie=TRUE,
vertex.cex=2)
text(Z, label=1:nrow(Z))
Three-dimensional Euclidean latent space model with three clusters and random
receiver effects
library(latentnet)
data(sampson)
monks.d3G3r <- ergmm(samplike~euclidean(d=3,G=3)+rreceiver)
plot(monks.d3G3r, rand.eff="receiver," use.rgl=TRUE,
labels=TRUE)
Bayesian ERGM
library(Bergm)
data(sampson)
mcmc<-bergm(samplike ~ edges + mutual + transitivities,
burn.in=20000, #Number of burn-in iterations
main.iter=100000, #Number of post-burn-in iterations
aux.iter=10000, #Auxiliary network: number of iterations
mprior=rep.int(0.0,3), #Prior: means of Gaussian
sdprior=rep.int(5.0,3), #Prior: standard deviations of Gaussian
nchains=1 #Number of Markov chains
)

```

REFERENCES

- Airoldi E, Blei D, Fienberg S, Xing E. Mixed Membership Stochastic Blockmodels. *Journal of Machine Learning Research*. 2008; 9:1981–2014. [PubMed: 21701698]
- Atchadé Y, Lartillot N, Robert CP. Bayesian Computing for Statistical Models with Intractable Normalizing Constants. *Brazilian Journal of Probability and Statistics*. 2012 to appear.

- Batagelj, V.; Mrvar, A. Pajek. Program for Large Network Analysis. Ljubljana, Slovenia: University of Ljubljana; 2003.
- Bernard HR, Killworth PD, Sailer L. Informant Accuracy in Social Network Data IV. *Social Networks*. 1979; 2:191–218.
- Besag J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society, Series B*. 1974; 36:192–225.
- Bhamidi, S.; Bresler, G.; Sly, A. Mixing Time of Exponential Random Graphs; 2008 IEEE 49th Annual IEEE Symposium on Foundations of Computer Science; 2008. p. 803-812.
- Brandes, U.; Lerner, J.; Snijders, TAB. Networks Evolving Step by Step: Statistical Analysis of Dyadic Event Data. *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*; IEEE; 2009. p. 200-205.
- Britton T, O'Neill PD. Statistical Inference for Stochastic Epidemics in Populations With Network Structure. *Scandinavian Journal of Statistics*. 2002; 29:375–390.
- Butts CT. Network Inference, Error, and Informant (In)Accuracy: A Bayesian Approach. *Social Networks*. 2003; 25:103–140.
- Butts CT. A Relational Event Framework for Social Action. *Sociological Methodology*. 2008; 38:155–200.
- Butts, CT. A Perfect Sampling Method for Exponential Random Graph Models. Irvine, CA: Department of Sociology, University of California; 2012.
- Caimo A, Friel N. Bayesian Inference for Exponential Random Graph Models. *Social Networks*. 2011; 33:41–55.
- Carrington, PJ.; Scott, J. The SAGE Handbook of Social Network Analysis. Vol. chap. 1. London: Sage; 2011. Introduction; p. 1-8.
- Celisse A, Daudin J-J, Pierre L. Consistency of Maximum-Likelihood and Variational Estimators in the Stochastic Block Model. 2011 Preprint available at <http://arxiv.org/pdf/1105.3288.pdf>.
- Daudin JJ, Picard F, Robin S. A Mixture Model for Random Graphs. *Statistics and Computing*. 2008; 18:173–183.
- Fellows, I.; Handcock, MS. Technical Report. Los Angeles: Department of Statistics, University of California; 2012. Exponential-Family Random Network Models.
- Frank O. Statistical Analysis of Change in Networks. *Statistica Neerlandica*. 1991; 45:283–293.
- Frank O, Strauss D. Markov Graphs. *Journal of the American Statistical Association*. 1986; 81:832–842.
- Geyer CJ. Likelihood Inference in Exponential Families and Directions of Recession. *Electronic Journal of Statistics*. 2009; 3:259–289.
- Geyer, CJ. R Package rcdd Version 1.1–7. Vienna, Austria: R Foundation for Statistical Computing; 2012.
- Geyer CJ, Thompson EA. Constrained Monte Carlo Maximum Likelihood for Dependent Data. *Journal of the Royal Statistical Society, Series B*. 1992; 54:657–699.
- Gile K, Handcock M. Respondent-Driven Sampling: An Assessment of Current Methodology. *Sociological Methodology*. 2010; 40:285–327. [PubMed: 22969167]
- Goldenberg A, Zheng A, Fienberg S, Airolidi E. A Survey of Statistical Network Models. *Foundations and Trends® in Machine Learning*. 2009; 2:129–233.
- Groendyke C, Welch D, Hunter DR. Bayesian Inference for Contact Networks Given Epidemic Data. *Scandinavian Journal of Statistics*. 2011; 38:600–616.
- Groendyke C, Welch D, Hunter DR. A Network-Based Analysis of the 1861 Hagelloch Measles Data. *Biometrics*. 2012; 68:755–765. [PubMed: 22364540]
- Häggström O, Jonasson J. Phase Transition in the the Random Triangle Model. *Journal of Applied Probability*. 1999; 36:1101–1115.
- Handcock, M. Technical Report. Seattle, WA: Center for Statistics and the Social Sciences, University of Washington; 2003. Assessing Degeneracy in Statistical Models of Social Networks. Available at <http://www.csss.washington.edu/Papers>.
- Handcock, MS.; Hunter, DR.; Butts, CT.; Goodreau, SM.; Krivitsky, PN.; Morris, M. *ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks*, Version 3.0–3. Seattle, WA: University of Washington; 2012. Available at <http://www.statnet.org>.

- Handcock MS, Raftery AE, Tantrum JM. Model-Based Clustering for Social Networks. *Journal of the Royal Statistical Society, Series A*. 2007; 170:301–354. (with discussion).
- Hanneke S, Fu W, Xing EP. Discrete Temporal Models of Social Networks. *Electronic Journal of Statistics*. 2010; 4:585–605.
- Hoff, P. Random Effects Models for Network Data. In: Breiger, R.; Carley, K.; Pattison, P., editors. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. Washington, DC: National Academies Press; 2003. p. 303–312.
- Hoff P. Bilinear Mixed-Effects Models for Dyadic Data. *Journal of the American Statistical Association*. 2005; 100:286–295.
- Hoff P. A Hierarchical Eigenmodel for Pooled Covariance Estimation. *Journal of the Royal Statistical Society, Series B*. 2009; 71:971–992.
- Hoff, P. Eigenmodel: Semiparametric Factor and Regression Models for Symmetric Relational Data, R Package Version 1.01. Vienna, Austria: R Foundation for Statistical Computing; 2012.
- Hoff PD, Raftery AE, Handcock MS. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*. 2002; 97:1090–1098.
- Holland PW, Leinhardt S. A Dynamic Model for Social Networks. *Journal of Mathematical Sociology*. 1977a; 5:5–20.
- Holland PW, Leinhardt S. Social Structure as a Network Process. *Zeitschrift für Soziologie*. 1977b; 6:386–402.
- Holland PW, Leinhardt S. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*. 1981; 76:33–65.
- Hummel RM, Hunter DR, Handcock MS. Improving Simulation-Based Algorithms for Fitting ERGMs. *Journal of Computational and Graphical Statistics*. 2012; 21:920–939.
- Hunter D, Lange K. A Tutorial on MM Algorithms. *The American Statistician*. 2004; 58:30–38.
- Hunter DR, Goodreau SM, Handcock MS. Goodness of Fit of Social Network Models. *Journal of the American Statistical Association*. 2008; 103:248–258.
- Hunter DR, Handcock MS. Inference in Curved Exponential Family Models for Networks. *Journal of Computational and Graphical Statistics*. 2006; 15:565–583.
- Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. *ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks*. *Journal of Statistical Software*. 2008; 24:1–29. [PubMed: 18612375]
- Jin, IH.; Liang, F. *Fitting Exponential Random Graph Models Using Stochastic Approximation MCMC*. Austin, Texas: Department of Biostatistics, University of Texas; 2012.
- Jonasson J. The Random Triangle Model. *Journal of Applied Probability*. 1999; 36:852–876.
- Kolaczyk, ED. *Statistical Analysis of Network Data: Methods and Models*. New York: Springer; 2009.
- Koskinen, J. *MelNet Social Networks Laboratory Technical Report 08-01*. Australia: Department of Psychology, School of Behavioral Science, University of Melbourne; 2009. The Linked Importance Sampler Auxiliary Variable Metropolis Hastings Algorithm for Distributions with Intractable Normalising Constants.
- Koskinen JH, Robins GL, Pattison PE. Analysing Exponential Random Graph (p-star) Models With Missing Data Using Bayesian Data Augmentation. *Statistical Methodology*. 2010; 7:366–384.
- Koskinen JH, Snijders TAB. Bayesian Inference for Dynamic Social Network Data. *Journal of Statistical Planning and Inference*. 2007; 137:3930–3938.
- Krivitsky PN. Exponential-Family Models for Valued Networks. *Electronic Journal of Statistics*. 2012; 6:1100–1128.
- Krivitsky PN, Handcock MS. Fitting Position Latent Cluster Models for Social Networks With Latentnet. *Journal of Statistical Software*. 2008; 24:1–23. [PubMed: 18612375]
- Krivitsky, PN.; Handcock, MS. Technical Report. University Park, PA: Department of Statistics, Pennsylvania State University; 2012. A Separable Model for Dynamic Networks.
- Krivitsky PN, Handcock MS, Raftery AE, Hoff P. Representing Degree Distributions, Clustering, and Homophily in Social Networks With Latent Cluster Random Effects Models. *Social Networks*. 2009; 31:204–213. [PubMed: 20191087]

- Leenders RTAJ. Models for Network Dynamics: A Markovian Framework. *Journal of Mathematical Sociology*. 1995; 20:1–21.
- Liang F. A Double Metropolis-Hastings Sampler for Spatial Models With Intractable Normalizing Constants. *Journal of Statistical Computing and Simulation*. 2010; 80:1007–1022.
- McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P., editor. *Frontiers in Econometrics*. New York: Academic Press; 1974. p. 105-142.
- McFadden D. A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*. 1989; 57:995–1026.
- Møller J, Pettitt AN, Reeves R, Berthelsen KK. An Efficient Markov Chain Monte Carlo Method for Distributions With Intractable Normalising Constants. *Biometrika*. 2006; 93:451–458.
- Morris M, Handcock MS, Hunter DR. Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *Journal of Statistical Software*. 2008; 24:1–24. [PubMed: 18612375]
- Murray, I. Ph.D. thesis. University College London: 2007. Advances in Markov Chain Monte Carlo Methods. Available at <http://www.cs.toronto.edu/murray/pub/>.
- Murray, I.; Ghahramani, Z.; MacKay, DJ. MCMC for Doubly-Intractable Distributions. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*; AUAI Press; Arlington, Virginia. 2006. p. 359-366.
- Nowicki K, Snijders TAB. Estimation and Prediction for Stochastic Blockstructures. *Journal of the American Statistical Association*. 2001; 96:1077–1087.
- Okabayashi S, Geyer CJ. Long Range Search for Maximum Likelihood in Exponential Families. *Electronic Journal of Statistics*. 2012; 6:123–147.
- Pattison P, Wasserman S. Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations. *British Journal of Mathematical and Statistical Psychology*. 1999; 52:169–193. [PubMed: 10613111]
- Perry PO, Wolfe PJ. Point Process Modeling for Directed Interaction Networks. 2011 under review. Available at <http://arxiv.org/abs/1011.1703>.
- Pflug, GC. *Optimization of Stochastic Models. The Interface Between Simulation and Optimization*. Boston: Kluwer Academic; 1996.
- Raftery AE, Niu X, Hoff PD, Yeung KY. Fast Inference for the Latent Space Network Model Using a Case-Control Approximate Likelihood. *Journal of Computational and Graphical Statistics*. 2012; 21:901–919.
- Rinaldo A, Fienberg SE, Zhou Y. On the Geometry of Discrete Exponential Families With Application to Exponential Random Graph Models. *Electronic Journal of Statistics*. 2009; 3:446–484.
- Ripley, R.; Snijders, TAB.; Preciado, P. *Manual for RSiena*. Department of Statistics, University of Oxford; 2011. Available at www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf.
- Robins G, Pattison P, Wasserman S. Logit Models and Logistic Regressions for Social Networks, III. Valued Relations. *Psychometrika*. 1999; 64:371–394.
- Robbins H, Monro S. A Stochastic Approximation Method. *Annals of Mathematical Statistics*. 1951; 22:400–407.
- Salter-Townshend M, Murphy TB. Variational Bayesian Inference for the Latent Position Cluster Model for Network Data. *Computational Statistics & Data Analysis*. 2013; 57:661–671.
- Sampson, S. Ph.D. thesis. Department of Sociology, Cornell University; 1968. A Novitiate in a Period of Change: An Experimental and Case Study of Relationships.
- Schweinberger M. Instability, Sensitivity, and Degeneracy of Discrete Exponential Families. *Journal of the American Statistical Association*. 2011; 106:1361–1370. [PubMed: 22844170]
- Schweinberger M. R Package Hergm Version 1.3–6. Hierarchical Exponential-Family Models for Networks. 2012a Available at <http://CRAN.R-project.org/package=hergm>.
- Schweinberger M. Statistical Modeling of Digraph Panel Data: Goodness-of-Fit. *British Journal of Mathematical and Statistical Psychology*. 2012b; 65:263–281. [PubMed: 21696381]
- Schweinberger, M.; Handcock, MS. Technical Report. Department of Statistics, Pennsylvania State University; 2011. Hierarchical Exponential-Family Random Graph Models: Local Versus Dependence.

- Schweinberger, M.; Petrescu-Prahova, M.; Vu, DQ. Technical Report. Department of Statistics, Pennsylvania State University; 2011. Disaster Response on September 11, 2001 Through the Lens of Statistical Network Analysis.
- Schweinberger, M.; Snijders, TAB. Settings in Social Networks: A Measurement Model. In: Stolzenberg, RM., editor. *Sociological Methodology*. Vol. vol. 33. Boston and Oxford: Basil Blackwell; 2003. p. 307-341.
- Schweinberger M, Snijders TAB. Markov Models for Digraph Panel Data: Monte Carlo-Based Derivative Estimation. *Computational Statistics & Data Analysis*. 2007; 51:4465–4483.
- Shalizi CR, Rinaldo A. Consistency Under Sampling of Exponential Random Graph Models. 2011 arXiv: 1111.3054v1.
- Shortreed S, Handcock M, Hoff P. Positional Estimation Within the Latent Space Model for Networks. *Methodology*. 2006; 2:24–33.
- Snijders, TAB. The Statistical Evaluation of Social Network Dynamics. In: Sobel, M.; Becker, M., editors. *Sociological Methodology*. Boston and London: Basil Blackwell; 2001. p. 361-395.
- Snijders TAB. Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure*. 2002; 3:1–40.
- Snijders TAB, Koskinen J, Schweinberger M. Maximum Likelihood Estimation for Social Network Dynamics. *Annals of Applied Statistics*. 2010; 4:567–588.
- Snijders TAB, Nowicki K. Estimation and Prediction for Stochastic Blockmodels for Graphs With Latent Block Structure. *Journal of Classification*. 1997; 14:75–100.
- Snijders TAB, Nowicki K. Manual for BLOCKS Version 1.8. 2007 Available at <http://www.gmw.rug.nl/stocnet/>.
- Snijders TAB, Pattison PE, Robins GL, Handcock MS. New Specifications for Exponential Random Graph Models. *Sociological Methodology*. 2006; 36:99–153.
- Snijders, TAB.; Steglich, CEG.; Schweinberger, M. Modeling the Co-Evolution of Networks and Behavior. In: van Montfort, K.; Oud, H.; Satorra, A., editors. *Longitudinal Models in the Behavioral and Related Sciences*. Mahwah, NJ: Lawrence Erlbaum; 2007. p. 41-71.
- Snijders, TAB.; Steglich, CEG.; Schweinberger, M.; Huisman, M. Manual for SIENA version 3.2. Department of Statistics, University of Oxford; 2012. Available at www.stats.ox.ac.uk/~snijders/siena/s_man317.pdf
- Snijders TAB, van de Bunt G, Steglich CEG. Introduction to Stochastic Actor-Based Models for Network Dynamics. *Social Networks*. 2010; 32:44–60.
- Stephens M. Dealing With Label-Switching in Mixture Models. *Journal of the Royal Statistical Society, Series B*. 2000; 62:795–809.
- Strauss D. On a General Class of Models for Interaction. *SIAM Review*. 1986; 28:513–527.
- Strauss D, Ikeda M. Pseudolikelihood Estimation for Social Networks. *Journal of the American Statistical Association*. 1990; 85:204–212.
- Tallberg C. A Bayesian Approach to Modeling Stochastic Blockstructures With Covariates. *Journal of Mathematical Sociology*. 2005; 29:1–23.
- van Duijn M, Gile K, Handcock M. A Framework for the Comparison of Maximum Pseudo Likelihood and Maximum Likelihood Estimation of Exponential Family Random Graph Models. *Social Networks*. 2009; 31:52–62. [PubMed: 23170041]
- van Duijn, MAJ. Estimation of a Random Effects Model for Directed Graphs. In: Snijders, TAB.; Engel, B.; Van Houwelingen, JC.; Keen, A.; Stemerding, GJ.; Verbeek, M., editors. *Toeval Zit Overal: Programmatuur Voor Random-Coëfficiënt Modellen*. Groningen, The Netherlands: IEC ProGAMMA; 1995. p. 113-131.
- van Duijn MAJ, Snijders TAB, Zijlstra BJH. P2: A Random Effects Model With Covariates for Directed Graphs. *Statistica Neerlandica*. 2004; 58:234–254.
- Vu, DQ.; Asuncion, AU.; Hunter, DR.; Smyth, P. Dynamic Egocentric Models for Citation Networks; *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*; 2011a. p. 857-864.

- Vu DQ, Asuncion AU, Hunter DR, Smyth P. Continuous-Time Regression Models for Longitudinal Networks. *Advances in Neural Information Processing Systems* 24 (NIPS 2011). 2011b:2492–2500.
- Vu, DQ.; Hunter, DR.; Schweinberger, M. Technical Report. Department of Statistics, Pennsylvania State University; 2012. Model-Based Clustering of Large Networks.
- Wang P, Robins G, Pattison P. PNet. Program for the Simulation and Estimation of Exponential Random Graph (p^*) Models. 2009 Available at <http://www.sna.unimelb.edu.au/pnet/pnet.html>.
- Wasserman, S. Ph.D. thesis. Department of Statistics, University of Harvard; 1977. Stochastic Models for Directed Graphs.
- Wasserman, S. A Stochastic Model for Directed Graphs With Transition Rates Determined by Reciprocity. In: Schuessler, KF., editor. *Sociological Methodology*. San Francisco, CA: Jossey-Bass; 1979. p. 392–412.
- Wasserman S. Analyzing Social Networks as Stochastic Processes. *Journal of the American Statistical Association*. 1980; 75:280–294.
- Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press; 1994.
- Wasserman S, Pattison P. Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^* . *Psychometrika*. 1996; 61:401–425.
- Westveld A, Hoff P. A Mixed Effects Model for Longitudinal Relational and Network Data, With Applications to International Trade and Conflict. *Annals of Applied Statistics*. 2011; 5:843–872.
- Wyatt, D.; Choudhury, T.; Bilmes, J. Learning Hidden Curved Exponential Random Graph Models to Infer Face-to-Face Interaction Networks From Situated Speech Data; *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*; 2008. p. 732–738.
- Zanghi H, Picard F, Miele V, Ambroise C. Strategies for Online Inference of Model-Based Clustering in Large and Growing Networks. *Annals of Applied Statistics*. 2010; 4:687–714.
- Zijlstra BJH, van Duijn MAJ, Snijders TAB. MCMC Estimation for the p_2 Network Regression Model With Crossed Random Effects. *British Journal of Mathematical and Statistical Psychology*. 2009; 62:143–166. [PubMed: 19208289]

Figure 1. Monk social network dataset of Sampson (1968), where polygons represent monks and directed edges represent liking relationships. The online version of this figure is in color.

Figure 2. Marginal posterior densities of canonical edges, mutual edges, and transitive edges parameters; the dotted lines indicate the MCMC MLE. Transitive edges should not be confused with the transitive triads of Table 1.

Figure 3. Posterior predictions of number of edges, mutual edges, and transitive edges; dotted line indicates observed value. Transitive edges should not be confused with the transitive triads of Table 1.

(a) Nowicki and Snijders (2001)

(b) Handcock et al. (2007)

(c) Krivitsky et al. (2009)

(d) Krivitsky et al. (2009)

Figure 4. Clustering of Sampson's monks by (a) stochastic block model of Nowicki and Snijders (2001); (b) latent space model of Handcock, Raftery, and Tantrum (2007); and latent space model of Krivitsky et al. (2009) in (c) two dimensions and (d) three dimensions. Loyal Opposition, Turks, and Outcasts are represented by red, green, and blue circles, respectively.

Table 1

Well-known network statistics, where circles represent nodes and directed lines represent directed edges between nodes. The first two of these statistics are used in the model of Section 4.4, whereas the last is replaced in the models by a statistic that is less prone to the degeneracy problem discussed in Section 4.3

	Edges	Mutual dyads	Transitive triads

$g(y)$	$\sum_{(i,j) \in \mathbb{Y}} y_{i,j}$	$\sum_{(i,j) \in u(\mathbb{Y})} y_{i,j} y_{j,i}$	$\sum_{(i,j) \in \mathbb{Y}} y_{i,j} \sum_{k \in N \setminus \{i,j\}} y_{i,k} y_{k,j}$
$\Delta_{i,j} g(y)$	1	$y_{j,i}$	$\sum_{k \in N \setminus \{i,j\}} (y_{i,k} y_{k,j} + y_{i,k} y_{j,k} + y_{k,i} y_{k,j})$

Table 2

Point estimates of edges, mutual edges, and transitive edges parameters, including posterior mean and standard deviation

	MPLE (SE)	MCMC MLE (SE)	Posterior mean (SD)
edges	-1.6981 (0.2502)	-1.9468 (0.3672)	-1.9216 (0.3535)
mutual edges	2.3277 (0.2940)	2.3093 (0.4195)	2.2732 (0.4273)
transitive edges	-0.0503 (0.1335)	0.1328 (0.2210)	0.1309 (0.2135)

NOTE: SE, standard error; SD, standard deviation.