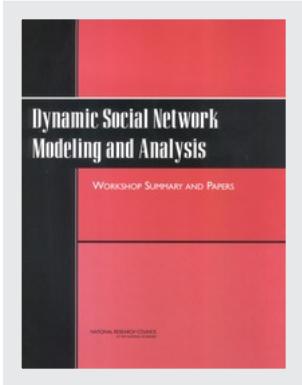


This PDF is available at <http://nap.nationalacademies.org/10735>



Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers (2003)

DETAILS

392 pages | 8.5 x 11 | PAPERBACK

ISBN 978-0-309-08952-4 | DOI 10.17226/10735

CONTRIBUTORS

Ronald Breiger, Kathleen Carley, and Philippa Pattison, Editors; Committee on Human Factors; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education; National Research Council

BUY THIS BOOK

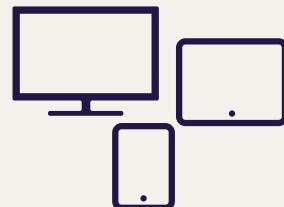
FIND RELATED TITLES

SUGGESTED CITATION

National Research Council 2003. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/10735>.

Visit the National Academies Press at nap.edu and login or register to get:

- Access to free PDF downloads of thousands of publications
- 10% off the price of print publications
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



All downloadable National Academies titles are free to be used for personal and/or non-commercial academic use. Users may also freely post links to our titles on this website; non-commercial academic users are encouraged to link to the version on this website rather than distribute a downloaded PDF to ensure that all users are accessing the latest authoritative version of the work. All other uses require written permission. ([Request Permission](#))

This PDF is protected by copyright and owned by the National Academy of Sciences; unless otherwise indicated, the National Academy of Sciences retains copyright to all materials in this PDF with all rights reserved.

Statistical Models for Social Networks: Inference and Degeneracy

Mark S. Handcock

Department of Statistics

Center for Statistics and the Social Sciences

University of Washington

Box 354322

Seattle, WA 98195-4322, USA

January 18, 2003

Abstract

This paper presents recent advances in the statistical modeling of random graphs that have an impact on the representation of social networks. We also consider issues related to the estimation of random graph models. For concreteness the focus is cross-sectional social networks.

Statistical exponential family models (Wasserman and Pattison, 1996) are a generalization of the Markov random graph models introduced by Frank and Strauss (1986), which in turn derived from developments in spatial statistics (Besag, 1974). These models recognize the complex dependencies within relational data structures.

To date, the use of stochastic graph models for networks has been limited by three interrelated factors: the complexity of realistic models, lack of use of simulation studies, and a poor understanding of the properties of inferential methods.

We discuss these factors and related issues of the degeneracy of inference for commonly promoted models. As a cornerstone of this development we present a Markov Chain Monte Carlo (MCMC) algorithm for general random graph models. We also review the role of these MCMC algorithms in simulation, likelihood-based inference, identifying degeneracy of inference, and Bayesian formulations.

KEYWORDS: Random graph models; Markov Chain Monte Carlo; Bayesian statistics.

1. Introduction

Networks are a form of “relational data”. Relational data arise in many social science fields and graph models are a natural approach to representing the structure of these relations. In these applications, the nodes usually represent people, and the edges represent a specified relationship between the people. This framework has many applications including, for example, the structure of social networks, the behavior of epidemics, the interconnectedness of the WWW, and long-distance telephone calling patterns.

We consider stochastic models for such graphs that form a statistical exponential family. This class has been referred to as the “ p^* ” class of models in the psychology and sociology literatures (Wasserman and Pattison, 1996). Given their general nature and applicability, we shall refer to them simply as random graph models. A much studied sub-class of models are the Markov random graph models introduced by Frank and Strauss (1986). These models

attempt to model the stochastic mechanism that produces the social ties and so represent the complex dependencies so induced. There is a large social network literature on this area which we do not review here. See Wasserman and Faust (1994), and Pattison and Robins (2000) for detailed information. The topic has connections to a broad array of literatures and here we emphasize the links to spatial statistics, statistical exponential families, log-linear models and statistical physics.

The exploration of the properties of graph models has been limited by three factors. First the complexity of realistic models has limited the insight that has been gained by analytical methods. Most analytical work has focused on simple one or two parameter models with independence between the dyads (Wasserman and Faust 1994, Frank 1997). Second statistical methods for the stochastic simulation from general random graph models have not existed. Because of this the properties of general models (e.g., range of graphs represented, dependence among the parameters) can not be explored through simulation studies. Third the properties of statistical methods for estimating the parameters based on observed networks are poorly understood (we consider inference in Section 3). Hence the range of parameter values relevant to real networks is largely unknown. In this paper we improve understanding of the nature and properties of graph models important to social networks by further considering methods for the stochastic simulation of, and inference for, random graphs.

We address one aspect of modeling that has been a persistent obstacle in the work in this area: *inferential degeneracy*. Many previous attempts to develop MCMC based estimation for Markov models have found that the algorithms nearly always converge to degenerate graphs – graphs that are either empty or full, or the algorithms do not converge consistently. Using statistical exponential family theory, we show that this is a function of the form of the model and algorithm used.

In the next section we review the statistical theory of social network models expressed in exponential form. We also consider MCMC algorithms to simulate graphs from these models. In Section 3 we consider forms of inference for the parameters of the models with focus on likelihood-based forms and Bayesian methods in particular. In Section we consider *model degeneracy* and some possible solutions to it.

2. Review of theory

Let the random matrix \mathbf{X} represent the incidence matrix of an undirected graph on n individuals. Thus \mathbf{X} is an $n \times n$ symmetric matrix, each of whose entries is a Bernoulli random variable; we assume further that the diagonal elements of \mathbf{X} are 0, which is to say that self-partnerships are disallowed. Suppose that \mathcal{X} denotes the set of all possible graphs on the given n individuals. The multivariate distribution of \mathbf{X} can be parameterized in the form:

$$P_\theta(\mathbf{X} = x) = \frac{\exp[\theta^T t(x)]}{c(\theta)} \quad x \in \mathcal{X} \quad (1)$$

where $\theta \in \Theta \subseteq \mathbb{R}^q$ is the model parameter and $t(x)$ is a q -vector of statistics based on the graph x . (Wasserman and Pattison, 1996). The denominator $c(\theta)$ is the constant that ensures the distribution sums to one: $c(\theta) = \sum_y \exp[\theta^T t(y)]$. Note that \mathcal{X} contains at most

$N = n(n-1)/2$ graphs, and $\Theta = \{\theta : c(\theta) < \infty\}$. The dimension of Θ is at most $2^N - 1$ (for the “saturated” model), although is typically much smaller than this. The parameter and

statistics that correspond to a given model for \mathbf{X} can be identified using the Hammersley-Clifford theorem (Frank and Strauss 1986). Conversely, each choice of parameter and graph statistics $t(x)$ specifies a model for \mathbf{X} . For example, if the dyads X_{ij} are mutually

independent the model can be written: $\log[P_\theta(X = x)] = \sum_{i < j} \theta_{ij} x_{ij} - \kappa(\theta)$, $x \in \mathcal{X}$ where

$\theta_{ij} = \text{logit}[P_\theta(X_{ij} = x_{ij})]$ and $\kappa(\theta) = \log[c(\theta)] = \sum_{i < j} \log[1 + \exp(\theta_{ij})]$. Thus $q=N$ and the

elements of $t(x)$ are just x_{ij} . This model is often called a Bernoulli graph. In the special case where the dyads have a common probability, $\log[P_\theta(X = x)] = \theta t(x) - \kappa(\theta)$ where $q=1$, $t(x)$ is the number of partnerships in the graph, and θ can be interpreted as the common log-odds of partnership formation within a dyad. The mathematical properties of this *homogeneous Bernoulli graph* model has been extensively studied (see, e.g., Renyi and Erdos 1959), although its simplicity and homogeneity make it less useful as a realistic model for social phenomena.

An alternative specification of the model (1) clarifies the interpretation of the parameters. Let $X_{ij}^c = \{X_{kl} : kl \neq ij, k < l\}$, $x_{ij}^c = \{x_{kl} : kl \neq ij, k < l\}$, $x_{ij}^+ = \{x_{ij}^c \text{ and } x_{ij} = 1\}$, and $x_{ij}^- = \{x_{ij}^c \text{ and } x_{ij} = 0\}$. Thus, X_{ij}^c represents all elements in the graph excluding X_{ij} , while x_{ij}^+ and x_{ij}^- represent the graph with the x_{ij} equal to 1 and 0, respectively. The full conditional distributions of X_{ij} are given by

$$\text{logit}[pr(X_{ij} = 1 | X_{ij}^c = x_{ij}^c)] = \theta^T \delta(x_{ij}^c) \quad x \in \mathcal{X} \quad (2)$$

where $\delta(x_{ij}^c) = t(x_{ij}^+) - t(x_{ij}^-)$ (Strauss and Ikeda 1991). The statistic $\delta(x_{ij}^c)$ is the change in the graph statistics when x_{ij} changes from 0 to 1. Hence θ can be interpreted as the increase in the conditional log-odds of a partnership between individuals i and j induced by the formation of the partnership and conditional on all other ties remaining unchanged. In the homogeneous Bernoulli graph, for example, θ is the common log-odds of individual partnership formation.

Holland and Leinhardt (1981) appear to be the first to propose log-linear models for social networks. Suppose that the dyads are independent with

$$\Pr(X_{ij} = x, X_{ji} = y) = \begin{cases} m_{ij} & \text{if } x = 1, y = l \\ a_{ij} & \text{if } x = 0, y = l \\ n_{ij} & \text{if } x = 0, y = 0 \end{cases} \quad (3)$$

Thus each dyad can have its own probability distribution. Thus the model can represent arbitrary attractiveness between individuals and degree of reciprocity within relationships. However the dyad independence implies specific transitivity and higher-order interactions. The model can be expressed in log linear form as:

$$\log[\Pr(X = x)] = \sum_{i < j} \rho_{ij} x_{ij} x_{ji} + \sum_{i \neq j} \theta_{ij} x_{ij} - \kappa(\theta) \quad x \in \mathcal{X}$$

They called this the p^1 model.

Based on developments in spatial statistics (Besag 1974), Frank and Strauss (1986) extended this work and introduced forms of dependence with Markov structure. Further extension were made by Wasserman and Pattison (1996) to incorporate actor attributes (Pattison and Wasserman 1999) and to allow explanatory and response variables (Robins, Pattison and Wasserman 1999), resulting in social influence (Robins, Pattison and Elliott 2000) and social selection models (Robins, Elliott and Pattison, 2000). These generalizations essentially allow

analysis of graphs with “colors” on the nodes, the coloring indicating the attribute. Recent developments have included new forms of dependency structures, to take into account social settings, and on the other hand a relaxation of Markovian dependence assumptions, allowing investigation of longer range configurations, such as longer paths in the network or larger cycles (Pattison and Robins, 2000). Models for bipartite (Faust and Skvoretz, 1999) and tripartite (Mische and Robins, 2000) graph structures have also been developed.

The incorporation of attribute information into the random graph model is straightforward. Suppose we wish to incorporate p exogenous covariates represented by a $n \times n \times p$ array of attributes W where the ijk^{th} element is the covariate for the p^{th} attribute on the ij^{th} dyad. Note that this allows the covariates to be attributes of the pairs of individuals (e.g., age difference) as well as specific to the individual alone. The graph statistic $t(x)$ in (1) is then replaced by $t(x, W)$, indicating that the statistics depend on the attribute information in addition to the relationship information. In this sense, the statistic $t(\cdot)$ can be defined to reflect the attribute information. We consider an implementation of this in the next section.

2.1. Cross-sectional models for random graphs

The most common form of random graph models exhibit Markov dependence in the sense of Frank and Strauss (1986). For these models, dyads that do not share an individual are conditional independent: an idea analogous to the nearest neighbor concept in spatial statistics. Typically a homogeneity condition is added: all isomorphic graphs have the same probability under the model. Frank and Strauss (1986) show that homogeneous Markov graphs are exactly those having the *triangle parameterization*: $q = n$, $\theta \in \Theta = \mathbb{I}^N$ and

$$t_k(x) = \frac{1}{k!} \sum_{i_0, \dots, i_k} x_{i_0 i_1} \dots x_{i_0 i_k} \quad k = 1, \dots, n-1 \quad t_n(x) = \frac{1}{6} \sum_{i,j,k} x_{ij} x_{jk} x_{ki},$$

where $t_k(x)$ is called a k -star and $t_n(x)$ is a count of the complete triads. An equivalent form is the degree distribution parameterization:

$$d_k(x) = \text{the proportion of individuals with exactly } k \text{ relationships} \quad k = 1, \dots, n-1$$

$$t_n(x) = \frac{1}{6} \sum_{i,j,k} x_{ij} x_{jk} x_{ki},$$

in which $d_k(x)$ counts the proportion of individuals with degree k . The degree distribution parameterization has the advantage that it is directly interpretable in terms of concurrency of partnerships (i.e. $d_m(x)$ for $m > 0$ counts the number of individuals with m concurrent partners).

A form of model that may realistically represent dependent social processes, reflect the impact of covariates is:

$$P(X = x) = \frac{\exp \left[x^T Z \beta + \sum_{k=0}^{n-1} \alpha_k d_k(x) + \theta^T t(x) \right]}{c(\alpha, \beta, \theta)} \quad x \in X \quad (4)$$

where x is the N -vector of the unique elements of X , $Z = \{z_{ij}\}_{N \times p}$ is a matrix of (exogenous) covariates on the ij^{th} dyad, $t(x)$ a q -vector of additional network statistics, β p -vector of regression parameters, α g -vector of degree parameters, and θ a q -vector of

network structure parameters.

This model is an extension of equation (1) and contains a number of special cases that have been considered in separate literatures. First the homogeneous Markov graphs of Frank and Strauss (1986) correspond to the case where there are no covariates and the additional network statistic is the count of complete triads. Second, if there are no covariates and no additional network statistics the model corresponds to the random graphs with arbitrary degree distributions considered in Newman, Strogatz and Watts (2001). This model assumes that all graphs with the same degree distribution are equally likely. On particular interest are the various “scale-free,” preferential attachment and power-law models popular in the physics literature (see, e.g., Newman 2002). Third, if there are no degree distribution parameters and no additional network statistics the model corresponds to the random graphs with independent dyads. The model still allows differential activity levels and assortative missing via the covariate parameters. Such models have been traditionally specified as log-linear models of mass action. See Morris, Goodreau and Koehly (2003) for an examination of the connections between the two.

As an example, consider a random graph model for a social network based on heterosexual partnerships. An example of a cross-sectional model that incorporates captures both concurrency and the assortative mixing based on race is:

$$P_\theta(X = x) = \exp\left[\sum_{k=1}^4 \alpha_k d_k(x) + \beta \text{race}(x)\right] / c(\alpha, \beta, \theta) \quad x \in X_h$$

$$P_\theta(X = x) = 0 \quad x \in X \setminus X_h$$

where $\text{race}(x)$ is the proportion of partnerships where both partners say they are of the same race, X_h is the subset of graphs in X in which all ties are heterosexual, and $\alpha \in \mathbb{R}^4$. The parameter α_1 represents the propensity to form monogamous relationships. The parameters $\alpha_2, \dots, \alpha_4$ represent the propensity to form concurrent partnerships with 2, 3 and 4 partners, respectively. Hence positive values for the latter parameters indicate the propensity of the graph toward higher degrees of concurrency. The parameter β represents the propensity for partnerships to be formed between individuals of the same race. Note that the graph and model reflect the heterosexual nature of the graph (i.e., $X_{ij} = 0$ unless i and j are of opposite sex). If $\alpha_1, \dots, \alpha_4$ are zero the dyads are independent and the model is simply assortative mixing by race. The model can then be re-written as a contingency table form (Morris, Goodreau and Koehly 2003). If β is zero then there are no preferences for partnerships based on race. We consider alternative parameterizations for this model in Section 3.

These models have a form that is flexible enough to capture the structure of many real networks. However to be practically useful we need to be able to simulate graphs with these distributions, infer their parameters from real data and measure their overall and relative goodness-of-fit. We now turn to each of these topics.

2.2. Simulating random graph models

The mathematical form of the models (1) and (4) allow graphs to be generated from them using Markov Chain Monte Carlo (MCMC) algorithms. Indeed MCMC algorithms have been much studied and are a natural way to simulate social networks (e.g. Gilks *et. al.* 1996,

Newman and Barkema 1999). The basic idea is to construct a Markov Chain on X with $P(X = x)$ as the equilibrium distribution. This is operationalized by starting from a graph in X , and making a large number of Markov transitions until approximate convergence to $P(X = x)$ is reached. Subsequent transitions are sampled and form a (dependent) sample from the desired model. For details, see the extensive literature cited in the above books. Many chains are possible with vastly different statistical properties. However convergence is “ensured” under mild conditions on the Markov Chain (irreducibility and aperiodicity). For the social network representation (1) this process has been studied by Crouch, Wasserman and Trachtenberg (1998), Coriander et. al. (1998), and Snijders (2002).

The full-conditional MCMC for the model (1) and (4) has a simple form and is motivated by the auto-logistic representation (2). In this algorithm each dyad is updated separately based on the conditional probabilities given in (2). This so-called “Gibbs sampling” or “heat bath” algorithm chooses the dyads randomly, sequentially, or using some mixture of the two. Each update requires the difference $\delta(x_{ij}^c) = t(x_{ij}^+) - t(x_{ij}^-)$ to be determined. The speed of the calculation of $\delta(x_{ij}^c)$ is an important factor in the computational quality of the algorithm (i.e., speed of convergence to $P(X = x)$, the relative frequency of producing each graph in X). Alternative Metropolis algorithms propose transitions from $x_{current}$ to $x_{proposed}$ based on the unconditional probabilities require the determination of $t(x_{proposed}) - t(x_{current})$. These tend to perform better for the same amount of computation. More general Metropolis-Hastings algorithms choose $x_{proposed}$ from an ancillary stochastic process dependent on $x_{current}$ and aimed at either focusing the transitions or spreading them more broadly through X . The behavior of these algorithms is very dependent on the choice of statistics $t(x)$.

The papers that use MCMC algorithms to simulate social network models report difficulties in convergence to realistic distributions (Crouch, Wasserman and Trachtenberg 1998, Coriander et al 1998, and Snijders 2002). A typical occurrence is for the algorithm to produce graphs that are complete, empty, or else are extreme in related ways. Coriander et. al. 1998 consider algorithms that condition on the number of ties in the graph. This ensure that the realizations are not complete or empty graphs. However, in most circumstances the density of the graph is a product of the social process that produced it and can not be assumed to be known in advance. Snijders 2002 reports on a number of related odd properties. In some cases the sequences of realizations transition quickly between very different graphs after periods of minor variation. In Section 3 we indicate why these algorithms behave in this way, how it relates to the model and how it can be resolved.

3. Inference for random graph models

Statistical methods for estimating the parameters in a random graph model are underdeveloped. Developing inference with a likelihood framework has the advantage of being able to draw upon a statistical theory for closely related models in statistical physics and spatial statistics (Besag 1975, Geyer 1999). This framework is also compatible with the Bayesian paradigm (Gelman, Carlin, Stern and Rubin 1995). The likelihood framework makes available exploratory graphical tools useful for inference about the underlying random field (Handcock, Meier and Nychka 1994). These tools, for example, can identify when a maximum likelihood approach is insufficient for inferential purposes.

Major barrier to the application of random graph models to social networks has been the lack of a sound statistical theory to evaluate how closely the models capture the structure in the observed graphs. This has two dimensions to it: the degree to which the graph structure of the models matched that in the data, and second the degree to which disease propagation through the model matches that of the data. The likelihood-based approach allow one to measure the first dimension of “goodness-of-fit” using Monte-Carlo p -value statistics and Bayes Factors under the Bayesian paradigm (Gelfand 1996).

3.1. Likelihood inference for random graph models

In principal, likelihood inference is straightforward for exponential random graph models. In this section we describe the most fundamental ideas and why direct computation is difficult. The log-likelihood for model (1) is:

$$L(\theta; x) \equiv \log[P_\theta(X = x)] = \theta^T t(x) - \kappa(\theta) \quad (5)$$

where $\kappa(\theta) = \log[c(\theta)]$. Likelihood inference is difficult as direct evaluation of $L(\theta; x)$ requires calculating

$$c(\theta) = \sum_{x \in X} \exp[\theta^T t(x)].$$

For even simple and modest models this can be problematic. Table 1 gives the number of elements in X (i.e., 2^N) as the number of actors grows.

Table 1: Number of Graphs for Different Numbers of Actors

Number of Actors, n	Number of Elements in X .
3	8
6	32,728
10	1.24×10^{27}
25	2.307×10^{90}

3.2. Alternative inference for random graph models

Clearly for realistic graphs direct enumeration is not feasible. Many alternatives have been proposed. Frank (1971) and Frank and Strauss (1986) consider (linear) approximations to the cumulant generating function:

$$C(\psi) = E[\log(\psi t(X))]$$

as a means to solve the *likelihood equations*

$$E_\theta[t(X)] = t(x_{\text{observed}}) \quad (6)$$

This approach is generally difficult to apply to general multiparameter models unless supplemented by a means of simulation from the same network model (Coriander et al 1998).

One commonly used method is *pseudolikelihood* originally motivated by, and developed for, spatial models by Besag (1975). The idea is to use an alternative local version of the

likelihood function defined the *pseudolikelihood*. The pseudolikelihood for model (1) is algebraically identical to the likelihood for a logistic regression of the unique elements of X on the design matrix with i th row $\delta(x_{ij}^c)$. The value of the maximum pseudolikelihood estimator for social networks can then be expediently found by using logistic regression as a computational device. Explicitly, the value of the maximum likelihood estimator for the logistic regression will also be the maximum pseudolikelihood estimator. Note, however, that the other characteristics of the maximum likelihood estimator do not necessarily carry over. In particular the standard errors of the estimates of θ from the logistic regression will not be appropriate for the maximum pseudolikelihood estimator. The statistical properties of pseudolikelihood estimators for social networks are only partially understood and are discussed in Handcock (2000). In the remainder of this paper we focus on likelihood based estimation.

3.3. Existence and uniqueness of MLE

The maximum likelihood estimator (MLE) for θ is:

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} L(\theta; x_{\text{observed}}) \quad (7)$$

Many properties of the MLE can be derived from statistical exponential family theory. Let C be the closed convex support of $t(X)$ - the convex hull of the discrete support points. Denote the interior of C by $\text{int}(C)$ and the boundary by ∂C .

Result:

- a) If Θ is open, the MLE exists if, and only if, $t(x_{\text{observed}})$ is inside the interior of C .
- b) If it exists, it is unique. In addition, when it exists, it can be found as the unique solution of (6) or equivalently as (7), the unique local minima of (5).
- c) As the support of $t(X)$ is discrete, and $t(x_{\text{observed}})$ is in it, a necessary and sufficient condition for the MLE not to exist is that $t(x_{\text{observed}})$ falls on the boundary of the support and that boundary be part of C .

The result is a fundamental property of statistical exponential families (Barndorff-Nielsen 1978).

In practice this means that attempt to numerically maximize the likelihood leads to unbounded estimates when the observed graph has statistics falling on the boundary of C . This typically means the optimization algorithm does not converge. Simulation studies in Handcock (2000) show that this is common for realistic models. If $t(x_{\text{observed}})$ falls on the boundary of C it is still possible that subsets of θ have MLEs that exist. Let $t(x_{\text{observed}}) = (t_a(x_{\text{observed}}), t_b(x_{\text{observed}}))$ and $\theta = (\theta_a, \theta_b)$ be similar partitions of $t(x_{\text{observed}})$ and θ . Under mild conditions, if $t_a(x_{\text{observed}})$ is on the boundary of C and $t_b(x_{\text{observed}})$ is in the relative interior of the convex hull of $\{t(X) : t_a(X) = t_a(x_{\text{observed}})\}$ then the MLE of θ_b exists and is the unique local minima of the conditional likelihood equation:

$$CL(\theta; x) \equiv \log [P_\theta(X = x | t_a(X) = t_a(x_{\text{observed}}))] = \theta^T t_b(x) - \kappa(\theta, t_a(x_{\text{observed}}))$$

$$x \in X : t_a(X) = t_a(x_{observed})$$

3.4. Likelihood inference based on MCMC algorithms

The difficulties in determining the normalizing function for the model (1) can be overcome using MCMC algorithms. The approach is based on the following idea: If a large number of simulations from a social network in the same ballpark as that of the observed network can be generated then these can be used to approximate the normalization function to a desired accuracy. Estimate the “population” value:

$$c(\theta) = \sum_{x \in X} \exp[\theta^T t(x)]$$

by the weighted “sample” mean:

$$\bar{c}(\theta) = |X| \sum_{M \text{ sampled graphs } y} \exp[\theta^T t(y)] w(y)$$

where the sample weights $w(y)$ are internally normalized to sum to unity. We can use the MCMC approach of Section 2.2 to simulate a sample of graphs from a model similar to the true, but unknown, model. The approximation to the likelihood using this approach is called the *MCMC likelihood*. The MLE for the data is then approximated by the maximum of the MCMC likelihood (the MC-MLE). This idea has been made precise and studied by Geyer and Thompson (1992). They show the MC-MLE converges to the true MLE as the number of simulations increases. The procedure also produces estimates of the asymptotic covariance matrix, the size of the MCMC induced error, and other related quantities.

As the sampled graphs form the basis of a statistical exponential family with sample space the realized values and probabilities the empirical proportions, the existence and uniqueness of the MC-MLE can be understood:

Result:

Let CO be the convex hull of *sampled* sufficient statistics. In practice, there are two situations:

- 1) $t(x_{observed}) \in int(CO) \subseteq int(C)$: The MC-MLE exists and is unique. It is found as the unique maximum of the MCMC likelihood.
- 2) $t(x_{observed}) \notin int(CO)$ but is in $int(C)$: The MC-MLE does not exist, even though MLE does.
- 3) $t(x_{observed}) \notin int(C)$: Neither the MC-MLE nor the MLE exists.

This result explains why attempts to calculate MC-MLE estimates for social network models fail. If the model used to simulate the graphs is not close enough to produce realizations that cover the observed values of the statistics, the MC-MLE will not exist even in cases where the MLE does. This behavior is quite common. As we shall see in Section 3, it is magnified by properties of commonly used models that do not place probability mass broadly enough. In sum, the MC-MLE may not exist for at least two reasons. First, the MLE itself may not exist (in which case neither does the MC-MLE). Second, it is difficult to specify parameter values for commonly used models to produce realizations that cover the observed values of the network statistics. In particular, simulations in Handcock (2000) show that simulations based on the maximum pseudolikelihood estimates usually do not produce realizations

sufficiently similar to those of the true values to support MC-MLE estimation. These properties of the models are very similar to those for spatial point processes considered in Geyer (1999).

3.5. Bayesian models and inference for random graph models

In random graph models of the form (1) the maximum likelihood estimator is not necessary optimal and, indeed, any point estimator of θ can be very poor. As such we need to move away from point estimators of θ and to frameworks for inference that naturally allow the uncertainty about the value of θ to be expressed.

We do this by supplementing likelihood-based inference with the Bayesian paradigm, mainly as it provides an elegant way of incorporating parameter uncertainty into the final inference and the incorporation of expert knowledge when it exists (Gelman, Carlin, Stern and Rubin 1995). Inference with the Bayesian paradigm, implemented via the now standard Markov Chain Monte Carlo (MCMC) methods can address, and even solve, many very difficult inferential problems, often making it the only realistic option.

Bayesian approaches to random graph models for social networks have received little attention (Wang and Wong 1987). We are in the process of developing Bayesian methodology for random graphs paying particular attention to the specification of prior distributions for the parameters that are meaningful for social networks. Under a Bayesian formulation, let $pr(\pi)$ be an arbitrary prior distribution for θ . The posterior distribution of θ is then $pr(\theta; X = x) = pr(\theta)P_\theta(X = x)/m(x)$ where $m(x) = \int_0^\infty pr(\phi)P_\theta(X = x)d\phi$ can be interpreted as the apriori probability that $X = x$. $pr(\theta; X = x)$ captures our knowledge about θ after the observed data (i.e., that is $X = x$) has been taken into account.

Note that $pr(\theta)$ may be chosen to restrict the parameter space to a subset of Θ by placing zero prior mass on the excluded values. Under these circumstances the posterior mass on these same values will always be zero regardless of their likelihood under the model and data. Computationally, Bayesian inference for an arbitrary prior is straightforward if MCMC simulation of the process is available. Conceptually this can be achieved by sampling a value θ^m from the prior distribution for θ , $pr(\theta)$, and then sampling \mathbf{X}^m from the distribution $P_{\theta^m}(X = x)$ using MCMC simulation. If this is repeated until \mathbf{X}^m equals the observed graph x_{obs} then the corresponding value of θ^m, θ^m (say), is a random sample from the posterior $pr(\theta; X = x_{obs})$. By repeating this process, a large sample of values of θ from $pr(\theta; X = x_{obs})$ can be generated. The posterior mean of θ can then be estimated by the mean of the sample. The posterior density corresponding to $pr(\theta; X = x_{obs})$ can be estimated by a non-parametric density estimate based on the sample. Note that this applies even when θ is multivariate, and so the posterior dependence between the components of θ can also be determined. There are alternative computational methods that are vastly superior in terms of computational efficiency to the above algorithm. These alternatives will almost always be used in practice. However, the above algorithm does show how Bayesian inference is achieved using MCMC simulation.

The definition and properties of families of conjugate priors for exponential families has been addressed in Barndorff-Nielsen (1978), Diaconis and Ylvisaker (1979) and, more recently,

Gutiérrez-Peña and Smith (1997). Of the many possible families possible, one of the most richest classes is discussed by Gutiérrez-Peña and Smith (1997). Let $g(\nu, \gamma)$ be the $(q+1)$ parameter exponential family with distributions:

$$\Pr(\theta; \nu, \gamma) = \frac{\exp[\nu^T \theta + \gamma \psi(\theta)]}{c(\gamma, \nu)} \quad \theta \in \Theta, \gamma \geq 0$$

Thus $g(\nu, \gamma)$ has canonical statistics θ and $-\psi(\theta)$ and is minimal unless $\psi(\cdot)$ is a linear function of θ . Here $\psi(\cdot)$ is a prespecified function. The *hyperparameters* ν and γ specify, respectively, the location and scale of θ . If $\Theta = \mathbb{R}^q$ then this distribution is proper if, and only if, ν/γ is in the interior of the convex hull of the sample space of $t(X)$ and $\gamma > 0$. Suppose the prior for θ is chosen to be $g(\nu, \gamma)$ then, formally, the posterior for θ , having observed $X = x$, is $g(\nu + t(x), \gamma + 1)$. That is,

$$\Pr(\theta; X = x; \nu, \gamma) = \frac{[(\nu + t(x))^T \theta + (\gamma + 1)\psi(\theta)]}{c(\gamma + 1, \nu + t(x))} \quad \theta \in \Theta, \gamma \geq 0$$

This family can be generalized to a p -variate γ . Let $g(\nu, \gamma)$ be the $(q+p)$ parameter exponential family with distributions:

$$\Pr(\theta; \nu, \gamma) = \frac{\exp[\nu^T \theta + \gamma^T \psi(\theta)]}{c(\gamma, \nu)} \quad \theta \in \Theta, \gamma_1 \geq 0, \dots, \gamma_p > 0$$

Here $\psi(\cdot)$ is a prespecified p -variate function. Thus $g(\nu, \gamma)$ has canonical statistics θ and $-\psi(\theta)$ and is minimal unless some component of the $\psi(\cdot)$ is a linear function of θ . Suppose the prior for θ is chosen to be $g(\nu, \gamma)$ then, formally, the posterior for θ , having observed $X = x$, is $g(\nu + t(x), \gamma + 1)$.

While inference for this family is quite tractable it is unclear under which circumstances the prior can be chosen to realistically represent knowledge about θ . It is more natural to choose the prior to reflect knowledge we do have and which is often quite strong. Examples include selectivity of graphs, degeneracy and stability. For these cases the prior family is unlikely to be conjugate.

4. Identifiability, Degeneracy and Stability for social networks models

The research reviews in the previous sections has allowed progress to be made in both the estimation and simulation of random graphs models. This work, and that of Besag (2000), has allowed provided insight into the properties of random graphs models. Two properties of random graph models that have a big impact on practice are *degeneracy* and *stability*. This builds on ideas of Ruelle (1969), Strauss (1986), Geyer (1999) and Baddeley (1999), Section 4. This work is in statistical physics and spatial point process theory.

The issue of what makes a useful model for a social network is a complex one. All models are representations of the manifestations of the social processes that produced the social network. Some models represent the proximate mechanisms of the formation of ties. This can be done by representation the temporal dynamics of the social process that produced the social network, or by direct representation of the network itself through appeal to concepts like structural balance and transitivity. Alternatively models may only aim at describing the probabilistic structure of the social ties. In all cases the models will be idealizations and simplification of the actual processes. Ultimately the answer to the question of what is a

useful model is fundamentally dependent on the purpose for which the model is intended. Simple models are often adequate, and even desirable, if there is a simple objective (e.g., what is the density of ties in the network). For other purposes, such as understanding the dynamics of the spread of STDs over networks, more sophisticated models are necessary.

Here we consider a basic property that a useful model should have. Broadly speaking, useful stochastic models place a significant proportion of their probability mass on graphs that have high probability of being produced by the underlying social process. We define a random graph model to be *degenerate* if the model places almost all its probability mass on a small number of graph configurations in X . Hence degeneracy of a graph model occurs when the model places disproportionate probability mass on only a few of the possible graph configurations. A common case is where the distribution places almost all its mass on the empty graph (i.e., $X_{ij} = 0 \quad \forall i, j$), and/or the complete graph (i.e., $X_{ij} = 1 \forall i, j$). Such models are not useful for modeling networks as almost all realizations from these models will be empty or full. In addition if such a model is used for simulation and MCMC likelihood inference the approximations to the true model will be very poor.

4. Discussion and future work

While estimation techniques are often of little interest to non-statisticians, this case is an exception. The complete enumeration of all graphs required by the denominator in (1) makes simple maximum likelihood estimation impossible for graphs larger than about 30 nodes. In early applications a form of pseudo-likelihood was used to solve this problem (Besag 1974, Wasserman and Pattison 1996). More recent approaches, however, employ Markov Chain Monte Carlo (MCMC) methods (Geyer and Thompson 1992). This is particularly interesting for our purposes because MCMC methods effectively simulate the network over the space of possible graphs in order to maximize the likelihood. One can, however, just as easily use the MCMC algorithm to simulate the network given the parameter estimates, and this provides the solution to the problem of linking network data to the network simulation.

This paper address the question of what is a good model and in doing so helps resolve the dilemma lack of convergence when estimation or simulating using MCMC. Many of the models that are proposed in literature suffer from degeneracy. Further algorithms used for inferential purposes are often inadvertently based on degenerate forms resulting in inferential degeneracy. One implication of these results is that the *effective* parameter space of exponentially parameterized random graph models is bounded and a small subset of the theoretical parameter space. The Bayesian framework for inference promises to be very powerful in social network modeling. It facilitates the propagation of parameter uncertainty into the final inference and allowing the incorporation of expert prior knowledge when it exists.