Proceedings of the 19th Annual Summer Political Methodology Meetings, 2002

Degeneracy and Inference for Social Network Models [1]

July 15, 2002

Abstract

The primary objective of this paper is to describe some recent advances in the modeling of random graphs that have an impact on the representation of sexual and drug use networks. We also investigate issues related to the estimation of random graph models. The main focus is cross-sectional social network models, although dynamic or longitudinal models are briefly discussed.

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.

In this paper we discuss these factors and related issues of the degeneracy of 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 for simulation, addressing model degeneracy, and likelihood-based inference.

KEY WORDS: Random graph models; log-linear network model; Markov fields; Markov Chain Monte Carlo.

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 discuss stochastic models for such graphs, with particular focus on sexual and drug use networks. The general class of models for such graphs 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), which in turn derived from developments in spatial statistics (Besag, 1974). These models recognize the complex dependencies within relational data structures.

There is a large literature on this area which we do not review here. See Wasserman and Faust (1994), and Pattison and Robins (2000) for detailed information. Throughput this paper I make connections to this literature, but also emphasize the links to spatial statistics, statistical exponential families, log-linear models and statistical physics.

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.

In this paper we discuss these factors and related issues of the degeneracy of 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 for simulation, maximum likelihood estimation, and Bayesian inference.

We also address one aspect of the problem that has been a persistent obstacle in the work in this area: *model 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. We show that this is a function of two things: models that are inappropriate for the data, and the general principle of selectivity bias. Most researchers select data for analysis that is non-degenerate, but the Markov model does not condition on this. Taking this into account in the estimation stage can have a dramatic impact on the estimates for models typically used in practice.

We also suggest alternative models that do not suffer from the degeneracy properties of

present models.

As part of this research program, we have produced a program for the estimation of random graphs. The program produces estimates of the random graph parameters using maximum psuedolikelihood estimates (MPLE) and MCMC maximum likelihood estimates (MLE) so that comparisons between the two can be made.

These ideas are applied to a sexual relations network from Colorado Springs with the objective of understanding the social determinants of HIV spread. The Colorado Springs Study (Klovdahl *et. al.*) is one of the first broad studies that attempts to understand the transmission of hepatitis B and HIV through a population based on a social network approach.

The study was developed by the Centers for Disease Control in Colorado Springs, CO and later funded by NIDA. A group of 595 persons at high risk for HIV infection were enrolled and interviewed for information concerning their medical and social background, their personal drug-using and sexual activity, and their social network contacts. These persons were sampled from street sites for prostitution, from the county methadone clinic, and from the county STD clinic. Persons named by several others were also actively recruited for the study, which qualifies this as a partial network sample design. As might be predicted from the focal nature of associations in this subculture, over 3600 of the 5812 contacts named by these persons were associated together in a single connected component (Woodhouse et al. 1994). A majority of persons with HIV were not in this large component, but were founded in smaller components. Those within the large components were demonstrated by a variety of measures to occupy non-central positions (Rothenberg et al. 1995a, 1995b). The authors concluded that in an area of low prevalence with lack of demonstrable endogenous transmission (Potterat et al. 1993), those with HIV were marginalized.

2. REVIEW OF THEORY

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 though simulation studies. Third the properties of statistical methods for estimating the parameters based on observed networks are poorly understood (we address inference in detail below). Hence the range of parameter values relevant to real networks is largely unknown. In this research we make major advances in understand the nature and properties of graph models important to risk networks by further improving methods we have developed for the stochastic simulation of and inference for random graphs.

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:

$$P_\theta(X = x) = \frac{\exp\left[\theta^T t(x)\right]}{\mathrm{c}(\theta)} \qquad x \in \mathcal{X} \tag{1}$$

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 $\mathrm{c}(\theta)$ is the constant that ensures the distribution sums to one: $\mathrm{c}(\theta) = \sum_{y\in} \exp\left[\theta^T t(y)\right] = 1$. 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 $\theta_{ij} = \mathrm{logit}[P_\theta(X_{ij} = x_{ij})]$ and $\kappa(\theta) = \log[\mathrm{c}(\theta)] = \sum_{i<j} \log[1 + \exp(\theta_{ij})]$. 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.

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

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

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. In the simple 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

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

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:

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

They called this the p^1 model.

Based on development 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 success of this modeling framework to risk behavior networks depends on properly specifying the relevant features, or configurations, of contact networks that can account for the global network properties. We focus on features from three domains: actor attributes, network patterns, and relationship timing.

Romantic relations tend to form within well defined attribute boundaries. Whether due to meeting opportunity (Feld 1981; McPherson and Smith-Lovin 1987) or preference (Billy, Rodgers, and Udry 1984; Davis 1963; Kandel 1978), similarity within relationship dyads is the norm. From a modeling standpoint, this implies that the probability of a partnership is much greater within same-attribute dyads than cross-attribute dyads. We focus on attributes such as race, socioeconomic status, geographic residence, occupation (sex-worker being a key distinction) and sex (capturing mainly sexual preference). Mixing preferences for type of sexual contact and IV drug-use are thus clearly important for modeling, and thus simulating, contact networks. Attribute mixing capture important nodal features of a sexual network's structure; nodes with certain attributes are likely to be connected. A model that stops there, however, must still assume random mixing within each attribute combination. The actual dynamics of disease transmission likely depend critically on the types of contact patterns beyond attribute or behavior constraints. To capture these features, we must identify how particular mixing patterns - beyond the type of partner – affect relationship formation. Patterns of interest are network stars of various sizes, paths, and cycles.

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) a 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 the attribute information defines the statistics.

The pattern of connections defines the set of all potential paths connecting actors in a sexual network. However, the timing of relations is important for determining which paths could potentially carry a pathogen. We focus on two aspects of relationship timing: concurrency and order. A pair of relations is concurrent when they share an actor in common at the same time. Thus, if i is connected to j, at time t and j is also connected to k at time t, then the ij and jk relations are concurrent. Past simulations have shown that networks with similar contact structure but differing levels of concurrency exhibit markedly different disease diffusion rates (Morris and Kretzschmar 1995; Morris and Kretzschmar 1997). Part

of the reason that concurrent relations pose a large infection risk lies in the way that indirect connectivity is channeled through concurrent relations. Two actors are indirectly connected if a sequence of sexual relations connects them through which a pathogen could travel. Since a pathogen can only travel forward in time, an indirect relation exists whenever a time-ordered path links two actors in the network. Thus, if i has sex with j, then later j has sex with k, k is at risk from i, but i is not at risk from k. When relations are concurrent, each actor linked through the concurrent relations can potentially infect every other actor, providing multiple, bi-directional indirect connectivity paths. While concurrency is the clearest feature of time-order in networks, indirect connectivity is conditioned by the order of relations more generally (Moody 2000). In networks where chains of relations unfold sequentially, indirect connectivity will flow asymmetrically. On the other hand, in chains that emerge as a result of multiple small chains merging at different times, indirect connectivity will be much lower. By modeling the timing as well as the pattern of sexual relations, we can account for indirect connectivity and thus disease diffusion risk.

We approach model development via a two step process. Initially, we focus on developing structural models for a given snapshot of the partnership network. These models account for the macro- and micro-level structure that we would expect to observe at a given point in time. Then, we add the temporal component to the model. We describe here our approach to modeling the sexual network in the Colorado Springs data, with similar approaches to be adopted to the needle-sharing network. The "static" epidemiological network models will account for nodal attribute mixing, concurrency, and degree mixing. The partnership networks that occur at a particular time point will be sparse, consisting of primarily isolates, stars, and paths. The nodes in these structures can be "colored" to account for the attribute mixing patterns that occur in the network. Using the modeling framework (1), Koehly and Morris (2000) defined an attribute mixing model for contact networks. They demonstrated that these attribute mixing models are directly related to the selective and proportional mixing models commonly used in the epidemiological network literature. The attribute mixing model will be extended to include multiple attributes (e.g. sex, ethnicity, and occupation), k-stars (concurrency), k-paths and path-star interactions (degree mixing).

2.1 Cross-sectional models for random graphs

The most common form of random graph model 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: $q = n, \theta \in \Theta = \mathbb{R}^n$ and

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

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:

$$d_k(x) = \begin{array}{c}\text{the number of individuals with} \\ \text{exactly} k \text{ relationships}\end{array} \qquad k = 1, ..., n-1 \qquad\qquad t_n(x) = \frac{1}{6} \sum_{i,j,k} x_{ij} x_{jk} x_{kl},$$

in which $d_k(x)$ counts the number of individuals with degree k. The degree distribution parameterization has the advantage of being based on statistics that are estimable from egocentric surveys. In addition they are interpretable in terms of concurrency of partnerships (i.e. $d_m(x)$ for $m > 1$ counts the number of individuals with m concurrent partners. As an example, consider a random graph model for a risk 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:

$$\log[P_\theta(X = x)] = \sum_{k=1}^{4} \theta_k d_k(x) + \theta_5 \text{race}(x) - \kappa(\theta) \qquad\qquad x \in \mathcal{X}$$

where race(x) is the number of partnerships where both partners say they are of the same race and $\theta \in \Theta = \mathbb{R}^5$. The parameter θ_1 represents the propensity to form monogamous relationships. The parameters $\theta_2, ..., \theta_4$ 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 θ_5 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 $\theta_1, ..., \theta_4$ 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 and Koehly 2000). If θ_5 is zero then there are not preferences for partnerships based on race. The goodness-of-fit of these interesting sub-models can be tested on using the methods developed here.

2.2 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.

The likelihood framework makes available exploratory graphical tools useful for inference about the underlying random field (Handcock, Meier and Nychka 1994). These tools can identify when a maximum likelihood approach is lacking.

As likelihood-based inference is based on $\mathcal{L}(\theta; X = x) \equiv \log[P_\theta(X = x)]$, it involves similar computational issues to those of stochastic simulation described above. In particular, direct calculation of $\mathcal{L}(\theta; X = x)$ by enumerating $\mathcal{X}$ is infeasible for the graphs of more than 30 individuals. Because of these computational difficulties, alternative means of estimation have been suggested. One is to use (linear) approximations to the cumulant generating function for $\mathbf{X}$ as a means to solve the likelihood equations (Frank 1971, Strauss 1986, Frank and Strauss 1986). 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 1997).

Major barrier to the application of random graph models to risk 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. Our likelihood-based approach allow us to measure the first dimension of "goodness-of-fit" using likelihood-ratio statistics and Bayes Factors under the Bayesian paradigm (Gelfand 1996).

2.3 Pseudo-likelihood inference for random graph models

Let $y = Vec(x)$, that is, writing the unique elements of the $n \times n$ symmetric matrix x in canonical form $y = \{x_{12}, x_{13}, \ldots, x_{1n}, x_{23}, \ldots, x_{2n}, x_{34}, \ldots, x_{(n-1)n}\}$. Thus y is a vector of length $N = n(n-1)/2$. Note that x can be reconstructed from y - it is just a way of re-writing it. In particular

$$\Pr(Y = Vec(x)) = \Pr(X = x) \qquad\qquad x \in \mathcal{X}$$

Let (ij) be the index of the element in $Vec(x)$ corresponding to the ij element in x. Formally $(ij) = j+n(i-1)-\frac{1}{2}i(i+1), i < j$ and, given an index k of y, $i = \min\{l : nl+\frac{1}{2}l(l+1) > k+1\}$, $j = k - n(i-1) + \frac{1}{2}i(i+1)$. Let $Y = Vec(X), X^c_{ij} = \{X_{kl} : kl \neq ij, kl \neq ji\}$, $x^c_{ij} = \{x_{kl} : kl \neq ij, kl \neq ji\}$, $Y^c_{(ij)} = Vec(X^c_{ij})$, and $y^c_{(ij)} = Vec(x^c_{ij})$. Denote by x^+_{ij} the sociomatrix with ijth element 1 and which agrees with x for all other dyads, that is, $x^+_{ij} = \{x^c_{ij}$ and $x_{ij} = 1\}$, Similarly define the sociomatrix like x, except with ijth element 0: $x^-_{ij} = \{x^c_{ij}$ and $x_{ij} = 0\}$. Then the full conditional distributions of the X_{ij} are specified by:

$$\text{logit}\left[\Pr(X_{ij} = 1 \mid X^c_{ij} = x^c_{ij})\right] = \text{logit}\left[\Pr(Y_{(ij)} = 1 \mid Y^c_{(ij)} = y^c_{(ij)})\right] = \theta^T \delta(y^c_{(ij)}) \qquad x \in \mathcal{X}$$

where

$$\delta(y^c_{(ij)}) = \delta(x^c_{ij}) = t(x^+_{ij}) - t(x^-_{ij})$$

Besag (1975) defined the *pseudolikelihood* for a parameter θ based on a finite set of random variables $Z_1, \ldots, Z_M$ as the product of the M individual full conditional distributions of each given the others. That is,

$$\mathrm{PL}(\theta; z) \equiv \prod_{i=1}^{m} P_\theta(Z_i = z_i | Z_j = z_j, j \neq i)$$

The pseudolikelihood estimator of θ is then

$$\tilde{\theta} = \mathrm{argmax}_{\theta \in \Theta} \mathrm{PL}(\theta; z)$$

The rationale for the use of pseudolikelihood as a surrogate is that if there is weak dependence among the variables $P_\theta(Z_i = z_i | Z_j = z_j, j \neq i) \approx P_\theta(Z_i = z_i)$ and $P_\theta(Z_1 = z_1, \ldots, Z_M = z_M) \approx \prod_{i=1}^{m} P_\theta(Z_i = z_i)$. Hence the pseudolikelihood and likelihood should be similar functions, and the pseudolikelihood estimator should have similar properties to the likelihood estimator.

Applying this idea to the social network:

$$\mathrm{PL}(\theta; y) \equiv \prod_{i=1}^{N} \Pr(Y_i = 1 \mid Y_i^c = y_i^c)^{y_i} \left[1 - \Pr(Y_i = 1 \mid Y_i^c = y_i^c)\right]^{1-y_i}$$

Note that if the dyads are independent, $\Pr(Y_i = 1 \mid Y_i^c = y_i^c) = \Pr(Y_i = 1)$ and the pseudolikelihood coincides with the likelihood. Thus when the graph is a (possibly inhomogeneous) completely random graph the two approaches are identical. In general, the dyads will be dependent and the two approaches will differ.

The value of the maximum pseudolikelihood estimator for social networks can be expediently found by using logistic regression as a computational device. Note that the pseudolikelihood (2.3) is algebraically identical to the likelihood for a logistic regression of Y on the design matrix with ith row $\delta(y_i^c)$. Thus 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. There appear to be three approaches to describing its properties.

The first, and most common, is to appeal to the similarity with the likelihood function. If the dyads have only *weak dependence* in the sense that $P_\theta(Z_i = z | Z_j = z_j, j \neq i) \approx P_\theta(Z_i = z)$, then the process is approximately inhomogeneously completely random network and the pseudolikelihood is an approximation to the likelihood. Hence the pseudolikelihood should have similar statistical properties to the likelihood. If the dyads are strongly dependent then

the approximation will be poor and the maximum pseudolikelihood estimator *may* perform poorly relative to the maximum likelihood estimator. This is a "hand-waving" argument that, to my knowledge, has not been quantified, and seldomly further justified.

The second approach is based on simulation studies of maximum pseudolikelihood on particular social networks. In these a test set of specific social network models are chosen *by design*, and the statistical properties of maximum pseudolikelihood described by calculating the estimates for many random draws from social networks. Example of such studies are Frank and Strauss (1986), Strauss and Ikeda (1990). However these studies have been of limited range both within model classes and in terms of the model classes considered. This has been partially due to two restrictions. First general software has not been available for the simulation of given social network models. An approach is given in Strauss (1986) and Frank and Strauss (1986). However this appears to sample from graphs conditional on the number of links, or where the number of links is additional chosen from a homogeneous completely random graph. Second the lack of general software to calculate the MLE has severely restricted comparisons of the MLE to the MPLE. As far as I know the small study in Strauss and Ikeda (1990) is the only example. Both these restrictions are removed with the availability of this software. Hence it is time for an extensive simulation study (see Section **??**).

The third approach is to describe various asymptotic statistical properties of the estimators. It is worthwhile understanding the statistical context within which these properties are studied. A primary role of asymptotic results is to provide approximations to the behavior of estimators for the sample sizes observable in practice. This is typically done by considering the behavior of the estimators as the number of observations increases. In statistical models where the observed data are independent, the corresponding maximum likelihood estimator can be shown to be consistent, efficient and asymptotically Gaussian. The efficiency is relative to the optimal estimator based on the observed data. These properties of the MLE have been extended to a broad range of statistical settings where the data are dependent. However in many statistical settings dependence exists between some or all of the observations. In these settings it is essential to specify ways for the number of observations to increase that are consistent with the nature of the datasets observed in practice. For example, consider a social network that follows a random graph model dependent on only the number of edges and 2-stars. One asymptotic framework is to consider the network as the number of actors increases. We can do this by considering independent replicates of a network of a fixed size precluding links between actors from different networks (e.g. friendship networks in schools from difference States). We can also do this by increasing the number of actors in a given network. However in this case the number of edges which each individual actor participates

in increases proportionately, and the number of 2-stars that an actor is involved in increases even faster. This can lead to unrealistic network realizations. An alternative is to limit the number or range of actors that an individual can relate to. A similar situation occurs in spatial statistics. Consider increasing the number of sites at which a random process is observed. One ways to do this is to let the observation region increase in such a way that the typical distance between neighboring sites remains constant ("increasing-domain asymptotics"). An alternative way is to keep the observation region fixed and let the typical distance between sites decrease ("fixed-domain" or "infill asymptotics"). Of course, here too there is an broad range of alternative schemes that can be motivated by considering how that site are chosen in practice. For a review of these issues see Ripley 1988, Baddeley and Gill 1997, Cressie 1993 and Stein 1999. In many settings where dependence exists, such as social networks, there are a range of possible ways to increase the number of actors, and these different choices often lead to different behaviors for the estimators. Each of these *asymptotic frameworks* leads to a different approximation to the properties of the estimator in practice. The accuracy and parsimony of these different approximations needs to be considered along with their motivation. For these reasons, little can be said in general about the behavior of estimators.

One closely related area which has received a lot of attention is that of discrete spatial Markov fields of regular lattices. In our context this is as if the actors are arranged in space on a regular grid (e.g. chess board) and only allowed to form relationship with their immediate spatial neighbors. See Guyon (1996), Baddeley (2000). Comets (1992) has proved strong consistency of the MPLE. Comets and Janžura (1998) show that the MPLE is asymptotic Gaussian. Additional description of the performance of the MPLE requires the model to satisfy regularity conditions that ensure uniqueness and stationarity of the limiting random field. Many of the concepts developed in this area are applicable to social networks. It is quite possible that these results can be modified so that they apply to social network models.

2.4 Markov Pseudo-likelihood inference for social networks

Perhaps the most common method of estimation for general graph models is maximum psuedo-likelihood. In the context of spatial Markov random fields, Besag (1986) has argued that the maximum pseudolikelihood estimator reflect the "local" (spatial) neighborhood information, as compared to the maximum likelihood estimator which reflects the "global" neighborhood information. I believe that the existing social network models are more global than local in structure, and that even the Markov forms of them are global. Hence my intuition is that the current approach to pseudolikelihood estimation will fail. The approach in this section is aimed at making the pseudolikelihood more global so as to partially overcome these defects.

Let $V = \text{Cov}(Y)$ be the covariance matrix of Y. Clearly the Y are independent only if X is a (possible inhomogeneous) completely random graph. In general V will not be a diagonal matrix, and the off-diagonal elements capture a component of the *dyad dependence.*

Let $\delta_{(ij)}$ be the indices of the Markov neighbors of the dyad (i, j) :

$$\delta_{ij} = \{(k, l) : k \in \{i, j\} \text{ or } l \in \{i, j\} \text{ but not both}\}$$

If X is a (homogeneous) Markov graph then, by symmetry,

$$V_{kl} = \begin{cases} \sigma^2 & \text{if } k = l \\ \sigma^2 \alpha_M & \text{if } k = \delta_l \\ \sigma^2 \alpha & \text{otherwise} \end{cases} \tag{4}$$

Thus σ^2 represents the variance of X_{ij}, α_M represents the correlation between dyads that are Markov neighbors (i.e., share an actor), and α represents the correlation between dyads that are not Markov neighbors. Note that σ^2 is a function of the (marginal) mean of X_{ij} as X_{ij} is Bernoulli.

It is important to note that these are the marginal covariances between the dyads rather than the covariance conditional on the rest of the network. As the values conditioned on are endogenous there will be a difference that should be adjusted for. The important of this observation still needs to be clarified, but note that the approach is still appropriately when the covariance structure is misspecified.

2.5 Likelihood inference for random graph models

In this section I describe the most fundamental ideas and why direct computation is difficult.

2.6 Simulation of random graphh models using MCMC procedures

Direct methods for simulating from random graph models are hampered by the numerical complexity of direct computation of the probabilities in (1). For example, a random graph of $n = 30$ individuals admits $2^{\binom{30}{2}} \approx 10^{131}$ possible graphs. Hence the calculation of $\text{c}(\theta)$ by direct enumeration of these is infeasible for graphs of most risk networks investigated in practice. It is now standard to simulate distributions of the form (1) by Markov Chain Monte Carlo (MCMC) methods (Gilks, Richardson, and Spiegelhalter 1996). The basic idea is to construct a discrete Markov chain on $\mathbf{X}$ with (1) as the equilibrium distribution. The chain is then run forward from some starting point until convergence and then sampled from. These samples are then are approximately random simulations from the random graph model. A plethora of such chains are possible and there is a vast literature on versions for specific multivariate distributions. These often have quite different computational and statistical properties (see, e.g., Gilks, Richardson, and Spiegelhalter 1996).

A Metropolis algorithm for random graphs is given in Strauss (1986) and Frank and Strauss (1986). However this applies to graphs conditional on the number of partnerships, or where the number of partnerships is additionally chosen from a homogeneous Bernoulli graph. Handcock (2000) gives an algorithm and software for simulation for a general random graph model based on sampling from the full conditional distributions (2). A similar algorithm is reported by Crouch, Wasserman and Trachtenberg (1997). We use this algorithm for simulation in the paper, and are developing more efficient and robust algorithms.

Recent advances by Propp and Wilson (1996) hold much promise. There approach, called *coupling from the past,* is a coupling algorithm for drawing exact samples from the equilibrium distribution of a discrete state Markov chain. By exact is meant that the samples have the equilibrium distribution rather than an approximation to the equilibrium distribution as is standard for the usual MCMC. These algorithms show great promise for social network models. In addition to automatically providing the correct distribution, they have been shown to be orders of magnitude faster than the usual MCMC algorithms for closely related spatial models.

Preliminary analysis of the structure of the correlations among the parameters show that they are highly correlated. Thus although the Gibbs sampler is very fast for each step it can take a long while to converge. It does, however, seem easy to measure via the usual diagnostics of Raftery and Lewis (1996) and Gelman (1996).

2.7 Likelihood inference for random graph models using MCMC

Geyer and Thompson (1992) show how MCMC methods can be used to compute the likelihood for a random graph model by estimating $\mathrm{c}(\theta)$ rather than calculating it exactly. The basic idea is that $\mathrm{c}(\theta)$ can be thought of as the mean of $\exp\left[\theta^T t(X)\right]$. One way determine to $\mathrm{c}(\theta)$ is to enumerate all graphs. As alternative is to randomly sample from the space of all graphs $\mathcal{X}$ and use the mean of the sample as an estimate of $\mathrm{c}(\theta)$. Geyer and Thompson (1992) do this with an importance weighted sampling scheme. The sampling can be achieved using a MCMC to simulate from the random graph model and the estimate can be made arbitrarily accurate by increasing the number of samples. It is then straightforward to calculate the maximum likelihood estimate, standard errors, and the size of the additional error incurred by the sampling scheme. We shall use the simulation method of Handcock (2000) in combination with the approach of Geyer and Thompson (1992) to construct maximum likelihood estimates of the models considered in this paper.

2.8 Identifiability, Degeneracy and Stability for social networks models

Recent work 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.

Degeneracy of a graph model occurs when the model places disproportionate probability mass on only a few of the possible graphs. The most common case is where the distribution places almost all its mass on the empty graph (i.e.,$X_{ij} = 0 \; \forall i, j$), and/or the complete graph (i.e.,$X_{ij} = 1 \; \forall i, j$). Such models are not useful for modeling risk networks as almost all realizations from these models will be empty or full. Handcock (2000) gives sufficient, and close to necessary, mathematical conditions on the model (1) for it to be degenerate. The degeneracy of a random graph is exacerbated as the number of individuals increase, but has a large effect on graphs of 7 or more individuals. He shows that many of the models that are proposed in literature suffer degeneracy and shows how to modify them to avoid the difficulties. One implication of these results is that the parameter space of non-degenerate models is bounded and a small subset of Θ. We are in the process of developing classes of models for risk networks that are non-degenerate. A random graph model is stable if small changes in the parameter values result in small changes in the probabilistic structure of the model. If this is not the case, then very similar parameter values can describe very different graph structures. Unstable models often have bad statistical properties are do not represent realistic graphs. Conditions for the stability of a model are related to those for degeneracy. We are developing classes of models for risk networks that are stable as well as non-degenerate.

In the context of spatial Markov random fields, Besag (1986) has argued that the maximum pseudolikelihood estimator reflects the "local" (spatial) neighborhood information, as compared to the maximum likelihood estimator which reflects the "global" neighborhood information. Results in Handcock (2000) suggests that commonly suggested random graph models are more global than local in structure, and that even those with Markov dependence are global. Hence the current approach to pseudolikelihood estimation will have poor statistical properties. This is supported by simulation results in Handcock (2000). This same global nature of the existing models contributes to their degeneracy and instability problems. These issues are not resolved by alternative forms of estimation but represent defects in the models themselves at least, to the extent that they are useful for modeling realistic graphs.

2.9 Local Models for social networks

Many of the problem with existing models is that they are too narrow and unrealistic to usefully describe social networks in practice. In addition they are globally defined which reduces their robustness and makes inference difficult. Pattison and Robins (2000) descibe a class of local random graph models that extend the theory presented here. However inference

and analysis can be done using the approaches presented here.

2.10 Longitudinal and Dynamic models for social networks

There is a growing need for graph models to represent changes in the graph over time (Doreian and Stokman 1997, Frank 1991, Suitor, Wellman and Morgan 1997). To capture the temporal dimension of the graph process, consider the stochastic process {**X** (t): 0≤ t≤ T} of n x n symmetric matrices where for given t,**X** (t) represents the graph of the risk network at time t. We shall refer to $\{\mathbf{X}(t) : t \geq 0\}$ as the dynamic graph. While general models for the dynamic graph can be fruitful explored we shall focus on continuous-time Markov models (Norris 1997, Snijders 2000). Consider the following model for the discrete dynamics of the graph from time t_1 to time $t_2 = t_1 + 1$:

$$P(X(t_2) = x(t_2) \mid X(t_1) = x(t_1)) = \frac{\exp\left[\phi^T s(x(t_2)) + \gamma^T q(x(t_1), x(t_2))\right]}{\mathrm{c}(\gamma, \phi, x(t_1))}$$
$$x(t_1), x(t_2) \in \mathcal{X}$$

This model has the same form as (1) and captures both the social and temporal aspects of the process. As in (1), the parameter ϕ and graph statistics $s()$ represent the underlying social structure. The parameter γ and statistics $q(x(t_1), x(t_2))$ represent the sociotemporal structure, that is, they identify how the structure of the graph at time t_1 influences it at time t_2. For example, consider the simple model with $q(x, y)$ equal to the number of dyads at time t_2 that are the same as at time t_1. Here γ expresses a tendency for partnerships (or non-partnerships) to persist from t_1 to t_2 . For most realistic dynamic processes, we would expect γ to be positive indicating the duration of ties overtime. Note that the temporal dependence under this model is determined by value of the same dyad at the two time points. In general, other aspects of the graph at time t_1 , such as the number of partners of the corresponding actors, may effect the value at time t_2 and correspond to alternative forms for $q(x, y)$. Although simple, this model of independence of separate dyads across time is the most commonly used in practice (Leenders 1997). Indeed with some notable exceptions (see, e.g., Doreian and Stokman 1997, Van de Bunt, Van Duijn and Snijders 1999, Robins and Pattison 2000), the issues of complicated dependence structures has receive little attention for models of dynamic random graphs. The issue here are the identification of realistic models for the change in a network over time. This is also often referred to as *network evolution*, *dynamic network models*, *temporal network processes*, or *longitudinal network processes*. A very good review of the state of the field in the middle 90's is the edited volume Doreian and Stokman (1997). It is required reading about the issues involved on longitudinal modeling. Unlike many of the other issues described in these notes, the modeling of the evolution of social networks is difficult and subtle.

There are many forms of network evolution over time. Doreian and Stokman (1997) describe four main classifications. Here we focus on model for the discrete transition of a network between two time points. This is akin to *event history analysis* of social networks. Among the many forms of evolving social networks, this is of particular interest as is fits in closely with the way that much social network data is collected through successive longitudinal surveys. However many of the other forms of evolution are also relevant to this case. For example, when the data is collected retrospectively at each point in the survey, a *continuous time* approach akin to *survival analysis* is possible. See Snijders (1996, 2001).

Here I summarize the approach of Robins and Pattison (2000). Random graph models can be used in this area by considering a *chain graph.* If we consider the transition of a network from point $t = 1$ to time $t = 2$, then we can consider a graph with the *parent* graph at time $t = 1$ and the *child* graph at time $t = 2$. Actors within each graph are connected with undirected links, while only directed links between parent and child are allowed. The models within each graph can take the usual random graph form (1). The issue is the specification of how the child graph links depend on the parent graph links. In principle these can be arbitrarily involved leading to complex dependencies. Robins and Pattison (2000) discuss this and propose various simplifications, including *complete perfect Markov models.*

This model can be used to simulate a longitudinal social network. The process is:

1. Generate a random graph model for the social network X^1 at time $t = 1$ using the MCMC described in the previous sections.

2. Use one of the random graph models in Robins and Pattison (2000) to generate a social network from $X^2 \mid X^1$, that is the social network at time $t = 2$ given that the network was X^1 at time $t = 1$. This can be achieved using a modified version of the MCMC algorithm that has already been developed. However, as the model *is* still of the form (1) it will require additional programming, but is straightforward otherwise.

3. Generate a social network from $X^3 \mid X^2$ using the same transition model as in the previous step. Repeat these steps for as many time points as are required.

Together the sequence of social networks $X^1, X^2, X^3, \ldots$ form a sample from a longitudinal social network. I believe the time evolution process should be kept very simple to start with. This will probably mean Markov dependence, where the *specific* dependence of $(ij)^2$ is only on $(ij)^1$, and its Markov neighbors ($(kl)^1$, where one of k or l is i or j), as well as its Markov neighbors at time $t = 2$, and global characteristics of X^1

It is worthwhile to keep separate the two forms of iteration involved in this process. The first is the *intrinsic time* step updating from $t = 1, 2, \ldots$ This refers to the actual

real-world temporal process. The second iteration is that of the MCMC updates. This is a *computational* step and will not correspond to the intrinsic time step, except as a coincidence. Fundamentally, it represents a computational device to obtain a sample from he correct distribution. Typically many (hundreds) of MCMC steps are required to produce a single sample from the intrinsic time update. Usually it is not even possible to interpret each MCMC step as a partial step in intrinsic time. In rare circumstances a MCMC algorithm could be constructed that has a direct correspondence with intrinsic time. However this is quite difficult to achieve for realistic models.

In practice this procedure can be used to simulate a continuous-time process by making the time step small and the number of time steps large. For example, we may record events to the closest week (i.e., which week in the year did you first have a relationship with ...). Then the stepping process be $t = 1, 2, \ldots$ weeks. The parameters of the model would need to be defined to reflect this - in particular, the probabilities of transition should be reduced accordingly if the time step is reduced.

2.11 Shortcomings of the random graph model

Although the random graph model (1) is conceptually simple, intuitively it seems problematic when we consider using it for social network data. The probability that a relationship exists between two individuals certainly depends on characteristics of the individuals, yet model (1) does not account for covariates and relies solely on summary statistics derived from the adjacency matrix X_{ij}.

Of course, models always fail to capture reality completely, so it is wise to ask whether model (1) is at least useful to some degree. Preliminary research suggests that they are not very useful, in the sense that quite often, most graphs simulated from distribution (1) are either complete or empty, which clearly does not reflect the reality we are interested in modeling.

Thus, these models need to be improved. One avenue of research is to investigate the properties of truncated versions of the random graph distribution—that is, distributions arising from constrained models. For example, we might consider exploring the properties of random graphs conditional on the total number of partnerships (or at least with the number of partnerships having a specified distribution). An additional computational challenge encountered in models of this type arises during the simulation phase, since a rejection sampling method that throws out all sampled graphs that don't satisfy the constraints is likely to be very inefficient due to the small proportion of graphs with distribution 1) that satisfy those constraints. We are developing efficient algorithms for simulating random graphs subject to constraints.

3. 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. One can estimate the network model from data, and then use the same model, with the empirically based parameter estimates, to drive a simulation of the network with an infection spreading through it. The MCMC algorithm provides the engine for both tasks. This will make it possible for the first time to directly control the network structures in a simulation so that they "look like" the networks we observe in different data sets. The simulation link provided by the MCMC algorithm will also make it possible to test how well the estimated models capture the network structure that matter for transmission. Both mixing and concurrency are examples of local organizing processes that affect reachability and distance in networks. One of the interesting questions from an applied perspective is whether the mixing and concurrency patterns largely determine the levels of reachability and distance. If so, then they provide a quick summary of the key network structures, and data collection can be organized accordingly. The MCMC simulation thus provides an opportunity to develop context sensitive "goodness-of-fit" tests for these random graph models. In sum, the data and models now exist to systematically address the key remaining questions in network epidemiology: how do networks matter for individual risk, and what impact do they have on the population dynamics of transmission.

References

Baddeley, A. J. (1999), "Likelihoods and Psuedolikelihoods for Markov Point Processes," in *Contributed paper for a festschrift for Professor Willem van Zwet*, Institute for Mathematical Statistics, p. in press.

— (2000), "Time-invariance estimating equations," *Bernoulli*, 6, 783–808.

Baddeley, A. J. and Gill, R. D. (1997), "Kaplan-Meier estimators of interpoint distance

distributions for spatial point processes," *Annals of Statistics*, 25, 263–292.

Bahadur, R. (1961), "A representation of the joint distribution of responses to n dichotomous items," in *Studies in Item Analysis and Prediction*, ed. Solomon, H., Stanford, California: Stanford University Press, pp. 158–168.

Barndorff-Nielsen, O. E. (1978), *Information and Exponential Families in Statistical Theory*, New York: Wiley.

Bearman, P. (1997), "Generalized exchange," *American Journal of Sociology*, 102, 1383–1415.

Bernardinelli, L. and Montomoli, C. (1992), "Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk," *Statistics in Medicine*, 11, 983–1007.

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series C*, 36, 192–236.

— (1975), "Statistical analysis of non-lattice data," *The Statistician*, 24, 179–95.

— (1986), "On the Statistical analysis of Dirty Pictures," *Journal of the Royal Statistical Society, Series C*, 48, 259–302.

— (1989a), "A Candidate's Formula: A Curious Result in Bayesian Prediction," *Biometrika*, 76, 183–183.

— (1989b), "Towards Bayesian Image Analysis," *Journal of Applied Statistics*, 16, 395–407.

— (2000), "Markov Chain Monte Carlo for Statistical Inference," Working paper, Center for Statistics and the Social Sciences, University of Washington.

Besag, J. and Higdon, D. (1999), "Bayesian Analysis of Agricultural Field Trials (with discussion)," *Journal of the Royal Statistical Society, Series C*, 61, 691–746.

Besag, J. and Kooperberg, C. (1995), "On Conditional and Intrinsic Autoregression," *Biometrika*, 82, 733–746.

Besag, J., York, J., and Mollièe, A. (1991), "Bayesian Image Restoration, with two applications is spatial statistics (with discussion)," *Annals of the Institute of Statistics and Mathematics*, 43, 1–59.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000), "Spatial Poisson regression for health and exposure data measured at disparate resolutions," *Journal of the American Statistical Association*, 45, 1076–1088.

Borg, I. and Groenen, P. (1997), *Modern Multidimensional Scaling: Theory and Applications*, Springer-Verlag.

Bradley Crouch, S. W. and Trachtenberg, F. (1998), "Markov Chain Monte Carlo Maximum Likelihood Estimation for p^* Social Network Models," in *Paper presented at the XVIII International Sunbelt Social Network Conference in Sitga, Spain.*

Comets, F. (1992), "On consistency of a class of estimators for exponential families of Markov random fields on the lattice," *Annals of Statistics*, 20, 455–468.

Comets, F. and Janžura, M. (1998), "A central limit theorem for conditionally centred random fields with an application to Markov fields," *Journal of Applied Probability*, 35, 608–621.

Cox, T. F. and Cox, M. A. A. (1994), *Multidimensional Scaling (ISBN 0412491206)*, Chapman & Hall.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

— (1995), "Bayesian smoothing of rates in small geographic areas," *Journal of Regional Science*, 35, 659–673.

Cwik, J. and Mielniczuk, J. (1989), "Estimating density ratios with application to discriminant analysis," *Communications in Statistics*, 18, 3057–3069.

— (1993), "Data-dependent bandwidth choice for a grade density kernel estimate," *Statistics and Probability Letters*, 16, 397–405.

DeSarbo, W. S., Kim, Y., and Fong, D. (1999), "A Bayesian Multidimensional Scaling Procedure for the Spatial Analysis of Revealed Choice Data," *Journal of Econometrics*, 89, 79–108.

Diaconnis, P. and Ylvisaker, D. (1979), "Conjugate Priors for Exponential Families," *Annals of Statistics*, 7, 269–281.

Doreian, P. and Stokman, F. N. (1997), *Evolution of Social Networks*, Amsterdam: Gordan and Breach.

Faust, K. and Romney, A. K. (1985), "Does Structure Find Structure?: A Critique of Burt's Use of Distance As a Measure of Structural Equivalence," *Social Networks*, 7, 77–103.

Fienberg, S. E., Meyer, M. M., and Wasserman, S. S. (1981), "Analysing Data From Multivariate Directed Graphs: An Application to Social Networks," in *Interpreting Multivariate Data*, Wiley (New York), pp. 289–306.

Fienberg, S. E. and Wasserman, S. S. (1981), "Categorical data analysis of single sociometric relations," *Sociological Methodology*, 156–192.

Frank, O. and Strauss, D. (1986), "Markov Graphs," *Journal of the American Statistical Association*, 81, 832–842.

Gelman, A. (1996), "Inference and Monitoring Convergence," in *Markov Chain Monte Carlo in Practice*, eds. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., New York: Chapman and Hall, pp. 131–144.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman and Hall.

Georgii, H. O. (1988), *Gibbs measures and phase transitions*, Berlin: De Gruyter.

Geyer, C. J. (1999), "Likelihood inference for spatial point processes," in *Stochastic Geometry: Likelihood and Computation*, eds. Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., Boca Raton: Chapman and Hall, pp. 79–140.

Geyer, C. J. and Thompson, E. A. (1992), "Constrained Monte Carlo maximum likelihood calculations (with discussion)," *Journal of the Royal Statistical Society, Series C*, 54, 657–699.

Guare, J. (1990), *Six Degrees of Separation: A Play*, Vintage: New York.

Gutiérrez-Peña, E. and Smith, A. F. M. (1997), "Exponential and Bayesian conjugate families: review and extensions," *Test*, 6, 1–90.

Guyon, X. (1996), *Random Fields on a Network: Modeling, Statistics and Applications*, New York: Springer.

Handcock, M. S. (1994), "Discussion of "Epidemics: Models and Data" by D. Mollison, V. Isham and B. Grenfell (1994)," *Journal of the Royal Statistical Society, Series A*, 157, 115–149.

— (1998), "Discussion of "Model–based Geostatistics" by P. J. Diggle, J. A. Tawn and R. A. Moyeed (1998)," *Journal of the Royal Statistical Society, Series C*, 47, 299–350.

— (2000), "Progress in Statistical Modeling of Drug User and Sexual Networks," Manuscript, Center for Statistics and the Social Sciences, University of Washington.

— (2001), "Degeneracy and Inference for Social Network Models," Working paper, Center for Statistics and the Social Sciences, University of Washington.

Handcock, M. S., Meier, K., and Nychka, D. (1994), "Comment on 'Kriging and Splines: An Empirical Comparison of their Predictive Performance' by G. M. Laslett," *Journal of the American Statistical Association*, 89, 401–403.

Handcock, M. S. and Morris, M. (1999), *Relative Distribution Methods in the Social Sciences*, New York: Springer.

Handcock, M. S. and Stein, M. L. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35, 403–410.

Handcock, M. S. and Wallis, J. (1994), "An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields (with discussion)," *Journal of the American Statistical Association*, 89, 368–378.

Hansell, S. (1984), "Cooperative groups, weak ties, and the integration of peer friendships," *Social Psychology Quarterly*, 47, 316–328.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983), "Stochastic Block Models: First Steps," *Social Networks*, 5, 109–137.

Holland, P. W. and Leinhardt, S. (1977), "A dynamic model for social networks," *Journal of Mathematical Sociology*, 5, 5–22.

— (1981), "An exponential family of probability distributions for directed graphs. With comments by Ronald L. Breiger, Stephen E. Fienberg, Stanley Wasserman, Ove Frank and Shelby J. Haberman and a reply by the authors," *Journal of the American Statistical Association*, 76, 33–65.

J., W. D. and H., S. S. (1998), "Collective dynamics of small-world networks," *Nature*, 393, 440–442.

J. Corander, K. D. and Dahmstr¨

om, P. (1998), "Maximum Likelihood Estimation for Markov Graphs," Research report, Department of Statistics, University of Stockholm.

Kelsall, J. E. and Wakefield, J. C. (1997), "Spatial Modeling of Disease Risk," Report, University of Washington.

Kent, D. (1978), *The rise of the Medici: Faction in Florence, 1426-1434*, Oxford University Press, Oxford.

Kretschmar, M. and Morris, M. (1996), "Measures of concurrency in networks and the spread of infectious disease," *Mathematical Biosciences*, 133, 165–195.

Lahiri, S., Kaiser, M., Cressie, N. A. C., and Hsu, N. (1999), "Prediction of Spatial Cumulative Distribution Functions Using Subsampling," *Journal of the American*

Statistical Association, 94, 100–102.

Lazarsfeld, P. F. and Henry, N. W. (1968), *Latent Structure Analysis*, Houghton Mifflin: Boston.

Milgram, S. (1967), "The small world problem," *Psychology Today*, 2, 60–67.

Moller, J. (1998), "Log Gaussian Cox Processes," *Scandinavian Journal Statistics*, 25, 451–482.

Moreno, J. L. (1934), "Who shall survive? A new approach to the problem of human interrelations," in *Psychological Abstracts*, Washington, DC: Nervous and Mental Disease Publishing Co, vol. 8, p. 5153.

Newman, M. E. J. (2000), "Models of the Small World: A Review," Tech. rep., World Wide Web.

Padgett, J. F. and Ansell, C, K. (1993), "Robust Action and the Rise of the Medici," *Amer. J. Sociology*, 98, 1259–1319.

Pattison, P. and Robins, G. (2000), "Neighbourhood-based models for social networks," Technical report, Department of Psychology, University of Melbourne.

Propp, J. G. and Wilson, D. B. (1996), "Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics," *Random Structures and Algorithms*, 9, 223–252.

Raftery, A. E. and Banfield, J. D. (1991), "Stopping the Gibbs Sampler, The Use of Morphology, and other Issues in Spatial Statistics," *Annals of the Institute of Statistics and Mathematics*, 43, 1–59.

Raftery, A. E. and Lewis, S. L. (1996), "Implementing MCMC," in *Markov Chain Monte Carlo in Practice*, eds. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., New York: Chapman and Hall, pp. 115–130.

Reitz, K. P. (1982), "Using Log Linear Analysis With Network Data: Another Look At Sampson's Monastery," *Social Networks*, 4, 243–256.

— (1988), "Social Groups in a Monastery," *Social Networks*, 10, 343–357.

Ripley, B. D. (1988), *Statistics Inference for Spatial Processes*, Cambridge: Cambridge University Press.

Robins, G. and Pattison, P. (2000), "p^* models for temporal processes in social networks," Technical report, Department of Psychology, University of Melbourne.

Ruelle, D. (1969), *Statistical Mechanics*, New York: Wiley.

S., O. M. and Raftery, A, E. (2001), "Bayesian multidimensional scaling and choice of dimension," *J. Amer. Statist. Association*, 96, 1031–1044.

S., S. J. and Tsai, C.-L. (1999), "Semiparametric and additive model selection using an improved Akaike information criterion," *Journal of Computational and Graphical Statistics*, 8, 22–40.

Sampson, S. F. (1968), "A novitiate in a period of change: An experimental and case study of relationships," Unpublished ph.d. dissertation, Department of Sociology, Cornell University.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, London: Chapman and Hall.

Sibson, R. (1979), "Studies in the Robustness of Multidimensional Scaling: Perturbational Analysis of Classical Scaling," *Journal of the Royal Statistical Society, Series B, Methodological*, 41, 217–229.

Snijders, T. A. B. (1996), "Stochastic Actor-Oriented Models for Network Change," *Journal of Mathematical Sociology*, 21, 149–172.

— (2001a), "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," Research report, Department of Statistics and Measurement Theory, University of Groningen.

— (2001b), "The Statistical Evaluation of Social Network Dynamics," *Sociological Methodology*, 31, to appear.

Snijders, T. A. B. and Nowicki, K. (2001), "Estimation and Prediction for Stochastic Block Structures," *Journal of the American Statistical Association*, 97, to appear.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer.

Strauss, D. (1986), "On a general class of models for interaction," *SIAM Review*, 28, 513–527.

Strauss, D. and Ikeda, M. (1990), "Pseudolikelihood estimation for social networks," *Journal of the American Statistical Association*, 85, 204–212.

Wang, Y. J. and Wong, G. Y. (1987), "Stochastic blockmodels for directed graphs," *Journal of the American Statistical Association*, 82, 8–19.

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

Wasserman, S. and Pattison, P. (1996), "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*," *Psychometrika*, 61, 401–425.

Wilson, D. B. (2000), "Layered Multishift Coupling for use in Perfect Sampling Algorithms (with a primer on CFTP)," *Fields Institute Communications*.

Zimmerman, D. L. and Cressie, N. A. C. (1992), "Mean squared prediction error in the spatial linear model with estimated covariance parameters," *Annals of the Institute of Statistics and Mathematics*, 44, 27–43.