Assessing Degeneracy in Statistical Models of Social Networks [1]

Mark S. Handcock
University of Washington, Seattle

Working Paper no. 39
Center for Statistics and the Social Sciences
University of Washington

December 31, 2003

[1]Mark S. Handcock is Professor of Statistics and Sociology, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4322. E-mail: handcock@stat.washington.edu; Web: www.stat.washington.edu/handcock. This research supported by Grant DA012831 from NIDA and Grant HD041877 from NICHD. The author is grateful to Carter Butts, Martina Morris, Pip Pattison, Garry Robins, Tom Snijders, Jim Moody, Julian Besag and Steve Goodreau for useful comments and discussions.

Abstract

This paper presents recent advances in the statistical modeling of random graphs that have an impact on the empirical study of 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 are derived from developments in spatial statistics (Besag 1974). These models recognize the complex dependencies within relational data structures. A major barrier to the application of random graph models to social networks has been the lack of a sound statistical theory to evaluate model fit. This problem has at least three aspects: the specification of realistic models, the algorithmic difficulties of the inferential methods, and the assessment of the degree to which the graph structure produced by the models matches that of the data. We discuss these and related issues of the model degeneracy and inferential degeneracy for commonly used estimators.

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

1. INTRODUCTION

Networks are a useful device to represent "relational data", that is, data with properties beyond the attributes of the individuals (nodes) involved. Relational data arise in many social science fields and graphical models are a natural approach to representing the regular pattern of the relations between nodes. In typical applications, the nodes in a graph represent individuals, and the ties (edges) represent a specified relationship between individuals. Nodes can also be used to represent larger social units (groups, families, organizations), objects (airports, servers, locations), or abstract entities (concepts, texts, tasks, random variables). This framework has many applications, including the structure of social networks, the dynamics of epidemics, the interconnectedness of the WWW, and long-distance telephone calling patterns.

We consider here stochastic models for such graphs. These models attempt to represent the stochastic mechanisms that produce ties, and the complex dependencies this induces. Collectively, the class of models we consider form a statistical exponential family (Strauss and Ikeda 1990). 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 exponentially parametrized random graph models, and note that in their general form they can represent any finite random graph model. There is a growing literature on these models – see Snijders (2002); Wasserman and Robins (2004). A natural sub-class of these are the Markov random graph models introduced by Frank and Strauss (1986). Exponentially parametrized random graph models have connections to a broad array of literatures in many fields, and here we emphasize its links to spatial statistics, statistical exponential families, log-linear models, and statistical physics.

Historically, exploration of the properties of these graphical models has been limited by three factors. First, the complexity of realistic models has limited the insight that can be obtained using analytical methods. Second, statistical methods for stochastic simulation from general random graph models have only recently been developed (Crouch et al. 1998; Corander and Dahmstrom 1998; Snijders 2002). Because of this, the properties of general models have not been explored in depth though simulation studies. Third, the properties of statistical methods for estimating model parameters based on observed networks have been poorly understood. The models and parameter values relevant to real networks is therefore largely unknown.

In this paper we examine the nature and properties of graphical models for social networks by extending methods for the stochastic simulation of, and inference for, random graphs. The approach builds on the work of Geyer and Thompson (1992).

Our main contribution is to address a persistent obstacle to applied work in this area: the problem of *inferential degeneracy*. Effective inferential strategies for general random graph models depend upon the use of Monte Carlo quadrature (Kalos and Whitlock, 1986) to estimate likelihood

values; efficient computation for such quadratures generally requires the use of Markov Chain Monte Carlo (MCMC) methods. Many previous attempts to develop MCMC-based estimation for Markov graph models have found that the algorithms nearly always converge to degenerate graphs – graphs that are either empty or complete – or that 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. In Section 3, we consider forms of inference for the parameters of the models with focus on psuedolikelihood and likelihood-based approaches. In Section 4 we review the geometry of random graph models, and in Section 5 use this to define a form of *model degeneracy*, a related form of inferential degeneracy, and some possible solutions. In Section 6 we illustrate these results using a simple random graph model that captures pair dependency.

2. RANDOM GRAPH MODELS

Let the random matrix $\mathbf{X}$ represent the adjacency matrix of an unvalued graph on n individuals. We assume that the diagonal elements of $\mathbf{X}$ are 0 – that self-partnerships are disallowed. Suppose that $\mathfrak{X}$ denotes the set of all possible graphs on the given n individuals. The multivariate distribution of $\mathbf{X}$ can be parameterized in the form:

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

where $\theta \in \Theta \subseteq \mathbb{R}^q$ is the model parameter and $t : \mathfrak{X} \to \mathbb{R}^q$ are statistics based on the adjacency matrix (Strauss and Ikeda 1990). There is an extensive literature on descriptive statistics for networks (Wasserman and Faust 1994; Borgatti et al. 1999). These statistics are often crafted to capture features of the network (e.g., transitivity, clustering, mutuality and betweenness) of primary substantive interest to the researcher. Often such substantive considerations cause the researcher to have a specific set of statistics in mind; the above model then has the property of maximizing the entropy within the family of all distributions with given expectation of $t(X)$. Paired with the flexibility of the choice of t this property does provide some justification for the model (1) that will vary from application to application.

The denominator $c(\theta, \mathfrak{X})$ is the *normalizing function* that ensures the distribution sums to one: $c(\theta, \mathfrak{X}) = \sum_{y \in \mathfrak{X}} \exp\left[\theta^T t(y)\right]$. This factor varies with both θ and the support $\mathfrak{X}$ and is the primary barrier to inference under this modeling scheme. We specify a model via the triple $(\mathfrak{X}, t, \theta)$.

Let C be the convex hull of $\{t(x) : x \in \mathfrak{X}\}$. Note that $\mathfrak{X}$ contains at most $N = 2^{n(n-1)}$ graphs so that the distributions in (1) form a statistical exponential model for the underlying finite exponential family of probability distributions with respect to counting measure on $\mathfrak{X}$. We assume

that the dimension of $\mathfrak{X}$ is q so that the family is minimal and the parameter space is $\{\theta : c(\theta, \mathfrak{X}) < \infty\} = \mathbb{R}^q$ so that it is also full. Although this is the most common situation for social network models, the results below hold with minimal changes if the parameter space is chosen to be a relatively open subset of $\mathbb{R}^q$. In the remainder we will suppress the reference to $\mathfrak{X}$. The dimension of Θ is at most $N - 1$ (for the "saturated" model), although it is typically much smaller than this.

An alternative specification of the model (1) clarifies the interpretation of the parameters. Let $\mathbf{I}$ be the set of indices of the unique elements of $\mathbf{X}$, $X_{ij}^c = \{X_{kl} : kl \in \mathbf{I}/\{ij\}\}$, $x_{ij}^c = \{x_{kl} : kl \in \mathbf{I}/\{ij\}\}$, $x_{ij}^+ = \{x_{ij}^c \cup \{x_{ij} = 1\}\}$, and $x_{ij}^- = \{x_{ij}^c \cup \{x_{ij} = 0\}\}$. Thus, X_{ij}^c represents all elements in the graph excluding X_{ij} while x_{ij}^+ and x_{ij}^- represent the graph x altered so that x_{ij} equal to 1 and 0, respectively. The full conditional distributions of X_{ij} are

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

where $\delta(x_{ij}^c) = t(x_{ij}^+) - t(x_{ij}^-)$ (Strauss and Ikeda 1990). 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 full conditional log-odds of a partnership between individuals i and j induced by the formation of the tie, conditional on all other ties remaining unchanged. In the homogeneous Bernoulli graph, for example, θ is the common log-odds of individual tie formation.

The most commonly used class of random graph models assume Markov dependence in the sense of Frank and Strauss (1986). For these models, dyads that do not share an individual are conditionally independent; this is an idea analogous to the nearest neighbor concept in spatial statistics. Typically a homogeneity condition is also added: all isomorphic graphs have the same probability under the model.

3. INFERENTIAL DEGENERACY FOR RANDOM GRAPH MODELS

Developing inference within 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 and Thompson 1992; Geyer 1999). Direct calculation of the log-likelihood:

$$\mathfrak{L}(\theta; x) \equiv \log[P_\theta(X = x)] = \theta^T t(x) - \kappa(\theta) \qquad x \in \mathfrak{X} \tag{3}$$

where $\kappa(\theta) = \log[c(\theta)]$, by enumerating $\mathfrak{X}$ is infeasible for graphs of more than 30 individuals. Because of these computational difficulties, alternative means of approximating the maximum likelihood estimator (MLE) for θ:

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} P_\theta(X = x_{observed} | t, \mathfrak{X}) \tag{4}$$

have been suggested. Frank (1971) and Frank and Strauss (1986) consider (linear) approximations to the cumulant generating function:

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

as a means to solve the likelihood equations

$$\mathbb{E}_\theta\left[t(X)\right] = t(x_{observed}) \tag{5}$$

where $\mathbb{E}_\theta(\cdot)$ is the expectation under P_θ. Unfortunately, this approach is generally difficult to apply to general multiparameter models unless supplemented by a means of simulation from the same network model (Corander and Dahmstrom 1998). Snijders (2002) considers stochastic approximation algorithms that attempt this. Until recently inference for the model (1) has be almost exclusively based on a local alternative to the likelihood function referred to as the *pseudolikelihood* (Strauss and Ikeda 1990). This was originally motivated by (and developed for) spatial models by Besag (1975). The pseudolikelihood for model (1) is algebraically identical to the likelihood for a logistic regression of the unique elements of the adjacency matrix on the design matrix with ith row $\delta(x_{ij}^c)$ (See (2)). The value of the maximum pseudolikelihood estimator can then be expediently found by using logistic regression as a computational device. Importantly, 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 (MPLE). While in common use (Wasserman and Pattison 1996; Anderson et al. 1999), the statistical properties of pseudolikelihood estimators for social networks are only partially understood.

3.1 Existence and uniqueness of MPLE

One concern with the maximum psuedolikelihood algorithm is that it produces infinite values even in situations where the psuedolikelihood converges. As it is usually assumed that all parameters are finite this is undesirable. If the MPLE does not exist then the estimates produced by software are determined by the precise convergence characteristics of the algorithm. As such, the estimates are arbitrary and misleading.

We now give a precise description of the existence and uniqueness of the MPLE for social network models.

Result:

1. A necessary and sufficient condition for the MPLE to exist (i.e., to be finite) is: $\forall \alpha \in \mathbb{R}^q, \exists i, j$ such that

   $$(2x_{ij} - 1)\alpha^T \delta(x_{ij}^c) \leq 0$$

   This occurs with positive probability.

2. If the MPLE exists, it is unique. In addition, when it exists, it can be found as the unique

solution to psuedolikelihood equations:

$$\sum_{i,j} x_{ij}\pi_{i,j} = \sum_{i,j} x_{ij}\delta(x_{ij}^c)$$

where $\text{logit}(\pi_{i,j}) = \theta^T\delta(x_{ij}^c)$.

The condition (1) has a simple interpretation: The MPLE exists unless a separating hyperplane exists between the scatterplot of the ties and non-ties in the space defined by the $\delta(x_{ij}^c)$. This behavior is referred to as "separation" when the model is logistic regression (Albert and Anderson 1984; Santner and Duffy 1986).

Heinze and Schemper (2002) suggest, in the context of logistic regression, a penalized likelihood approach. For social networks, we can consider a penalized psuedolikelihood using Jeffreys invariant prior as the penalty function (Heinze and Schemper 2002; Firth 1993). Under this modification the estimates will be finite and may reduce the bias of order $O(N^{-1})$.

3.2 Existence and uniqueness of MLE

Many properties of the MLE can be derived from statistical exponential family theory (Barndorff-Nielsen 1978). Denote the relative interior of C by rint(C) and the relative boundary by $\text{rbd}(C) = \text{cl}(C)\backslash\text{rint(C)}$.

Result:

1. The MLE exists if, and only if, $t(x_{observed}) \in \text{rint(C)}$
2. If it exists, it is unique. In addition, when it exists, it can be found as the unique solution to (5) or, equivalently, as (4), the unique local minima of (3).
3. A necessary and sufficient condition for the MLE not to exist is that $t(x_{observed}) \in \text{rbd}(C)$. This occurs with positive probability.

In practice, the above implies that attempting to numerically maximize the likelihood leads to unbounded estimates when the observed graph has statistics falling on $\text{rbd}(C)$. This typically means the optimization algorithm does not converge, or otherwise converges to a false maxima. Simulation studies by Handcock (2000), Snijders (2002); Snijders et al. (2004) show that this is a common occurrence as many realistic models have a non-negligible probability of falling on $\text{rbd}(C)$.

If $t(x_{observed})$ falls in $\text{rbd}(C)$ it is still possible that subsets of θ have MLEs that exist. Let $t(x_{observed}) = (t^{(1)}(x_{observed}), t^{(2)}(x_{observed}))$ and $\theta = (\theta^{(1)}, \theta^{(2)})$ be similar partitions of $t(x_{observed})$ and θ. Under mild conditions (Barndorff-Nielsen 1978), if $t^{(1)}(x_{observed}) \in \text{rbd}(C)$ and $t^{(2)}(x_{observed})$ is in the relative interior of the convex hull of $t(X) : t^{(1)}(X) = t^{(1)}(x_{observed})$ then the MLE of $\theta^{(2)}$ exists and is the unique local minima of the conditional likelihood equation:

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

$$x \in \mathfrak{X} : t_a(X) = t_a(x_{observed}) \tag{6}$$

Thus, it may be possible to obtain estimates for certain model parameters despite difficulties with the overall model. This may provide a fall-back for certain atypical scenarios in which models fail despite being well-behaved in general (e.g., due to chance realizations of extreme values of $t(x_{observed})$ in small graphs).

3.3 Likelihood-based inference based on MCMC Algorithms

Geyer and Thompson (1992) show how MCMC methods can be used to approximate the likelihood for the random graph model (1). The basic idea is that $c(\theta, \mathfrak{X})$ can be thought of as the mean of $\exp\left[\theta^T t(X)\right]$. One way determine to $c(\theta, \mathfrak{X})$ is to enumerate all graphs in $\mathfrak{X}$. As is well-known, an alternative is to randomly sample from $\mathfrak{X}$ and use the (weighted) mean of the sample as an estimate of $c(\theta, \mathfrak{X})$: Let $y^{(1)}, \ldots, y^{(M)}$ be a sequence of graphs sampled from the model $(\mathfrak{X}, t, \psi)$. A natural Monte Carlo estimate of $c(\theta, \mathfrak{X})$ is

$$c_M(\theta, \mathfrak{X}) = \frac{c(\psi, \mathfrak{X})}{M} \sum_{i=1}^{M} e^{(\theta-\psi)^T t(y^{(i)})}$$

and the corresponding *Monte Carlo log-likelihood*:

$$\mathrm{MC}\mathfrak{L}(\theta; x, \psi) = (\theta - \psi)^T t(x) + \log \left[\sum_{i=1}^{M} e^{(\theta-\psi)^T t(y^{(i)})} \right] + \log M$$

converges almost surely to the (relative) log-likelihood $\mathfrak{L}(\theta; x) - \mathfrak{L}(\psi; x) = (\theta - \psi)^T t(x) + \kappa(\psi, \mathfrak{X}) - \kappa(\theta, \mathfrak{X})$. The MLE of θ can then be approximated by: $\hat{\theta}_{MC} = \mathrm{argmax}_{\theta \in \Theta} \mathrm{MC}\mathfrak{L}(\theta; x_{observed}, \psi)$, the *Monte Carlo MLE* (MC-MLE). Geyer and Thompson (1992) prove that 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.

The existence and uniqueness of the MC-MLE can be understood in terms of the statistical exponential family with respect to counting measure on $\{t(y^{(1)}), \ldots, t(y^{(M)})\}$.

Result: Let CO be the convex hull of $\{t(y^{(1)}), \ldots, t(y^{(M)})\}$. There are three situations:

1. If $t(x_{observed}) \in$ rint(CO) the MC-MLE exists and is unique. It is found as the unique maximum of the MCMC likelihood.

2. If $t(x_{observed}) \in$ rint(C) $\cap$ (rint(CO))^c then the MC-MLE does not exist, even though MLE exists and is unique.

3. If $t(x_{observed}) \notin$ rint(C) then neither the MC-MLE nor the MLE exists.

This result further clarifies why attempts to calculate MC-MLE estimates for social network models often 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 (Crouch et al. 1998; Snijders 2002). As we shall see in Section 5, this effect is magnified by properties of commonly used models that do not place probability mass broadly enough. Thus, 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 – when the importance sample fails to cover these values, the MC-MLE will not exist. This second problem is complementary to the Monte Carlo estimation error: small to moderate disparities between $P_\psi(X = y)$ and $P_{\hat{\theta}}(X = y)$ lead to inaccurate estimation, with larger disparities causing the complete failure of the algorithm.

4. GEOMETRY OF EXPONENTIALLY PARAMETRIZED RANDOM GRAPH MODELS

The exponentially parametrized random graph models have almost exclusively been analyzed within the natural $\theta-$parametrization. In this section we establish the geometry of two alternative parameterizations of model (1) that are usually superior for social networks.

Consider the mean value parametrization for the model $(\mathfrak{X}, t, \theta)$, $\mu : \Theta \to$ rint(C) defined by

$$\mu(\theta) = \mathbb{E}_\theta\left[t(X)\right] \tag{7}$$

The mapping is strictly increasing in the sense that

$$(\theta_a - \theta_b)^T(\mu(\theta_a) - \mu(\theta_b)) \geq 0 \tag{8}$$

with equality only if $P_{\theta_a}(X = x) = P_{\theta_b}(X = x)\ \forall x$. The mapping is also injective in the sense that

$$\mu(\theta_a) = \mu(\theta_b) \to P_{\theta_a}(X = x) = P_{\theta_b}(X = x)\ \ \forall x. \tag{9}$$

Using (1), $\mu(\theta)$ can be reexpressed as

$$\mu(\theta) = \left[\frac{\partial \kappa(\theta)}{\partial \theta}\right](\theta) \tag{10}$$

and has gradient

$$\nu(\theta) = \left[\frac{\partial^2 \kappa(\theta)}{\partial \theta^T \partial \theta}\right](\theta) \tag{11}$$

The range of μ, $\mu(\Theta)$, is rint(C) (Barndorff-Nielsen 1978). Taken together, this means that (7) is an alternative parameterization of $(\mathfrak{X}, t, \theta)$. We refer to it as the $\mu-$parametization.

The $\mu-$parametization has many advantages. Chief among these is interpretability. As noted in Section 2, researchers choose graph statistics $t(x)$ because they have meaning in the substantive context. The parameter corresponding to each graph statistic is just its expected value over the population of graphs. For many parameters, this population-average interpretation has great intuitive appeal. The $\mu-$parameter space is finite and convex, and the points in the space have an interpretable scale.

A natural measure of the divergence of the model $(\mathfrak{X}, t, \mu_0)$ from $(\mathfrak{X}, t, \mu)$ is the Kullback-Leibler discriminant information:

$$\mathrm{KL}(\mu_0; \mu) = \sum_{x \in \mathfrak{X}} \log\left[P_{\mu,\mathfrak{X}}(X = x)/P_{\mu_0,\mathfrak{X}}(X = x)\right] P_{\mu,\mathfrak{X}}(X = x) = (\theta(\mu)-\theta(\mu_0))^T \mu + \kappa(\mu_0) - \kappa(\mu)$$

As

$$\frac{\partial \mathrm{KL}(\mu_0; \mu)}{\partial \mu_0} = \nu^{-1}(\theta(\mu))(\mu_0 - \mu) \qquad \frac{\partial \mathrm{KL}(\theta_0; \theta)}{\partial \theta_0} = \nu(\theta)(\theta_0 - \theta) \tag{12}$$

we see that small changes on the $\mu-$parametrization may lead to large-scale divergence while those in the $\theta-$parametrization are bounded. These issues are illustrated in the next sections.

In the case where a sub-set of the statistics are most directly interpretable in terms of their mean values and others in their natural parametrization we can consider mixed parametizations. Let $(t^{(1)}, t^{(2)})$ be a partition of t such that the first component is that of the statistics interpretable in terms of their mean values. Consider similar partitions $(\theta^{(1)}, \theta^{(2)})$ of θ and $(\mu^{(1)}(\theta), \mu^{(2)}(\theta))$ of $\mu(\theta)$. Let $\Theta^{(2)}$ be the set of values of $\theta^{(2)}$ for θ varying in Θ and $C^{(1)}$ be the convex hull of $\{t^{(1)}(x) : x \in \mathfrak{X}\}$. The mapping $\eta : \Theta \to \Theta^{(2)} \times \mathrm{rint}(\mathrm{C}^{(1)})$ defined by

$$\eta(\theta) = (\mu^{(1)}(\theta), \theta^{(2)}) \tag{13}$$

represents a third parametization of the model $(\mathfrak{X}, t, \theta)$ (Barndorff-Nielsen 1978). The parametrization has similar properties to the mean value parametrization. In addition it has the important property that the components $\mu^{(1)}$ and $\theta^{(2)}$ are variationally independent, that is, the range of $\eta(\theta)$ is a product space. We refer to it as the $\eta-$parametization.

For many random graph models the $\eta-$parametization has the advantages of interpretability and better inferential properties than the natural or mean value parameterizations.

4.1 Estimation in General Parametizations

Let ξ be one of the three parameterizations of the model $(\mathfrak{X}, t, \theta)$. Let $\kappa(\xi) = \log[c(\theta(\xi))]$ where $\theta(\xi)$ is the inverse mapping of $\xi(\theta)$. The log-likelihood function:

$$\mathfrak{L}(\xi; x) \equiv \log\left[P_{\theta(\xi)}(X = x)\right] = \theta(\xi)^T t(x) - \kappa(\xi) \qquad x \in \mathfrak{X} \tag{14}$$

is continuous in ξ and satisfies

$$\frac{\partial^2 \mathfrak{L}(\xi; x)}{\partial \xi^T \partial \xi} = \frac{\partial \theta}{\partial \xi^T} V_{\xi^T} \frac{\partial \theta}{\partial \xi} + (t(x) - \mathbb{E}_\xi\left[t(X)\right]) \cdot \frac{\partial^2 \theta}{\partial \xi^T \partial \xi} \tag{15}$$

where V_{ξ^T} is the variance of $t(X)$ considered as a function of ξ. In addition, if $t \in \text{rint}(C)$ the observed information at the MLE $\hat{\xi}$ is

$$-\frac{\partial^2 \mathfrak{L}(\xi; x)}{\partial \xi^T \partial \xi}(\hat{\xi}) = \mathfrak{I}(\hat{\xi}) = \mathbb{V}_{\hat{\xi}}\left[t(X)\right] \tag{16}$$

where $\mathfrak{I}(\xi)$ is the Fisher information. The Fisher information for the $\theta-$parametrization is $\nu(\theta) = \mathbb{V}_\theta\left[t(X)\right]$ while that of the $\mu-$parametrization is $\nu^{-1}(\theta)$.

The Fisher information of the $\eta-$parametrization has additional structure. Based on the identity:

$$\frac{\partial^2 \mathfrak{L}(\eta; x)}{\partial \mu^{(1)T} \partial \theta^{(2)}}(\eta) = (t^{(1)}(x) - \mu^{(1)}) \cdot \frac{\partial^2 \theta^{(1)}}{\partial \mu^{(1)T} \partial \theta^{(2)}} \tag{17}$$

the information is

$$\begin{bmatrix} [\nu^{-1}(\theta)]_{11} & 0 \\ 0 & [\nu(\theta)]_{22} \end{bmatrix}$$

where the diagonal blocks are the corresponding components of the information for $\mu^{(1)}$ and $\theta^{(2)}$, respectively. Thus the components of the $\eta-$parametrization are orthogonal, and the corresponding MLEs are asymptotically independent. These two properties enhance the interpretation of the $\eta-$parametrization and simplify inference.

It is possible to enlarge the model (1) to include the distributions corresponding to the faces of C (Barndorff-Nielsen 1978, p.154-155). In this extended model, the MLE of μ, $\hat{\mu} = t(x_{observed})$, always exists and is unique. While no direct calculation of the likelihood equation is necessary to calculate the MLE, measuring the uncertainty of the MLE as an estimate of μ is as computationally complex as finding the MLE of θ.

The MLE of η can be determined in two stages. The MLE of $\mu^{(1)}$ is clearly $t^{(1)}(x_{observed})$. The MLE of $\theta^{(2)}$ can be found from the conditional likelihood equation (6). If this is estimated based on an MCMC then in general it will be simpler than an overall MCMC as the proposal function can be restricted to graphs with the observed value of $t^{(1)}$.

For all three parameterizations, an MCMC or analytic approximation is required to determine $\nu(\hat{\theta})$.

These results provide insight into the failure of the MLE algorithms in the natural parametrization noted in Section (3.3).

5. MODEL DEGENERACY FOR RANDOM GRAPH MODELS

In this section we consider the relationship between the geometry of the parametrization and the properties of the model. Two properties of random graph models that have a big impact on practice are *model degeneracy* and *stability.* This builds on ideas of Ruelle (1969), Strauss (1986), Geyer (1999) and Baddeley (2001, Section 4).

5.1 What makes a useful model for a social network?

The specifics of what makes a useful model for a social network depends on the application. Here we consider a basic statistical property that any 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. Heuristically, model degeneracy occurs when the model places disproportionate probability mass on only a few of the possible graph configurations in $\mathfrak{X}$. A common case is where the random graph model $(\mathfrak{X}, t, \theta)$ 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 almost never useful for modeling actual networks, as almost all realizations from these models will be empty or complete. While it may be true that some social structures *are* nearly always empty or complete this invariance also means that these are rarely the subject of network analysis. Another complication caused by model degeneracy is that when such models are used for simulation and MC-likelihood inference, the approximations to the true model will generally be very poor. Thus, for most applications, we seek $(\mathfrak{X}, t, \theta)$ which are far from degenerate and we seek to limit our space of viable models accordingly.

Specifically, we say a model $(\mathfrak{X}, t, \theta)$ is *near degenerate* if $\mu(\theta)$ is close to the boundary of the convex hull of C. Let $\deg(\mathfrak{X}) = \{x \in \mathfrak{X} : t(x) \in \text{rbd}(C)\}$ be the set of graph on the boundary of the convex hull. Based on the geometry of the mean value parametrization this means that the expected sufficient statistics are close to a boundary of the hull and the model will place much probability mass on graphs in $\deg(\mathfrak{X})$. This statement can be quantified in a number of ways:

Result: Let e be a unit vector in $\mathbb{R}^q$ and $\text{bd}(e) = \sup_{\mu \in \text{rint(C)}}(e^T \mu)$.

1. $\mu(\lambda e) \to \text{bd}(e)e$ as $\lambda \uparrow \infty$.
2. $P_{\lambda e, \mathfrak{X}}(X \in \deg(\mathfrak{X})) \to 1$ as $\lambda \uparrow \infty$.
3. For every $d < \text{bd}(e)$, $P_{\lambda e, \mathfrak{X}}(e^T t(X) \leq d) \to 0$ as $\lambda \uparrow \infty$.
4. Let $\theta_0 \in \text{rint(C)}$. Then $\text{KL}(\theta_0; \lambda e) \to \infty$ as $\lambda \uparrow \infty$.

As $\mu(\theta)$ approaches the boundary the corresponding model places more and more probability mass on the graphs that form the corresponding part of the boundary. The limiting model is singular with respect to models with $\mu(\theta) \in \text{rint(C)}$ and places positive probability on the small subset of graphs in the corresponding face of C. The last result indicates that near degenerate models diverge in the Kullback-Leibler sense from models which are not.

In practice, the proportion of the population of graphs in $\deg(\mathfrak{X})$ is small. Often the models place the majority of mass on the empty graph, the complete graph or a mixture of the two.

These issues are exacerbated in the natural parametization due to the complexity of the mapping $\mu(\theta)$. Typically only a small section of the natural parameter space is mapped to areas of rint(C) far

from the boundary. As a result it is difficult to specify models in the natural parameter space. The set of realistic social network models often form one-dimensional curves in the natural parameter space of models that are surrounded by near degenerate models. We illustrate these issues using the $2-$star model in the next section.

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 do not represent realistic graphs. Conditions for the stability of a model are related to those for model degeneracy. The stability of model within the mean value parametrization is geometrically simpler than that of the natural parametrization: if the model is unstable if it is close to rbd(C).

5.2 The Effect of Model Degeneracy on MCMC Estimation

The effects of model degeneracy are exacerbated because it is closely related to poor properties of simple MCMC schemes (Geyer 1999). The results of the previous section along with those of Section 3 clarify why attempts to calculate MC-MLE estimates for social network models often fail. Parameter values in the near degenerate region can hinder the convergence of common MCMC algorithms used to simulate and estimate random graph models. This has hindered likelihood-based inference for random graph models. A common symptom of these difficulties has been non-convergence of the algorithms (Crouch et al. 1998; Snijders 2002). Specifically, if the MCMC is based on a near degenerate model then the convex hull of the sampled statistics will be a small subset of the convex hull of the model. According to the results of Section 3.3, the MC-MLE will have very high variance and usually will not exist. The reason is that it is very difficult to choose values of the natural parameter that are far from degenerate and close to the MLE.

Consider the simple MCMC algorithm based on the full-conditional distributions and the random proposal of dyads to update. Let

$$M(\psi) = \max_{x \in \mathfrak{X}} |\psi^T \delta(x_{ij}^c)|$$

where $\delta(x_{ij}^c)$ is given in (2). Thus there exists $x \in \mathfrak{X}$ with .

$$\text{logit}\left[P_{\psi,\mathfrak{X}}(X_{ij} = 1 \mid X_{ij}^c = x_{ij}^c)\right] = \pm M(\psi)$$

If $M(\psi)$ is large and the chain transitions to this graph then it will not mix over $\mathfrak{X}$ very well. This is true even though the MCMC is geometrically ergodic. However, as $\mu(\psi) \to \text{rbd}(C)$, $M(\psi) \to \infty$. Hence if ψ is near-degenerate then $M(\psi)$ will often be large and the will have poor coverage of $\mathfrak{X}$. A common occurrence with social network models has $M(\psi)$ determined by the complete graph, and the MCMC unable to move away from it.

6. CASE-STUDY OF DEGENERACY OF THE 2−STAR MODEL

In this section we illustrate the ideas in the previous section using a simple model that is commonly used as a starting point for social network analysis. It is the instance of the model (1) with two network statistics being the number of edges and 2−stars in the graph. This is the case of model (1) with $q = 2$ statistics

$$P_\theta(X = x) = \frac{\exp\{\theta_1 E(x) + \theta_2 S(x)\}}{c(\theta_1, \theta_2)} \qquad x \in \mathfrak{X}$$

$$E(x) \equiv t_1(x) = \sum_{i<j} x_{ij}$$

$$S(x) \equiv t_2(x) = \sum_{i<j<k} x_{ij} x_{ik}$$

where $E(x)$ is the number of edges and $S(x)$ is the number of "2−stars" (i.e., edges sharing a common node). We shall refer to this as the *2-star* model. While it only has two parameters it can represent the full degree distribution of a graph. This model modifies the homogeneous Bernoulli model to capture a naive form of dependence between pairs measured by $S(x)$. It is an analog of the Strauss model for a point process (Strauss 1975) or the Ising model for a lattice process.

The natural parameter space of models is $\theta \in \Theta \subseteq \mathbb{R}^2$; that is, any value of θ in the plane corresponds to a valid probability distribution. In this illustration we will consider the case where there are $g = 7$ homogeneous actors. For this model there are $N = 21$ pairs and $|\mathfrak{X}| = 2{,}097{,}152$ graphs. The degeneracy of this random graph model grows as the number of individuals increase, but its effects can be seen even in graphs as small as 7 actors.

Figure 1 is a plot of the sufficient statistics for the model. Each circle represents a pair of sufficient statistics (i.e., numbers of edges and 2−stars). In total there are 144 distinct pairs of sufficient statistics. The area of each circle represents the number of isometric graphs with that pair of sufficient statistics. For example, for the pair $(10, 22)$ there are 79,170 graphs and for $(10, 23)$ there are 55,230 graphs.

The convex hull of the support C is plotted. On the boundary of C there are 22 pairs of statistics, all of them on the bottom of the hull. These correspond to graphs with the minimum possible number of 2−stars for the given number of edges. From the earlier results, if the observed statistics is one of these 22, then the MLE of θ_2 does not exist, and the MLE of θ_1 is the log-odds of the observed density. If the observed statistic is not one of these then the MLE can be calculated from (4).

We are less concerned with estimation here, and more concerned with the nature of the probability mass functions defined by the 2−star model. What do the graphs generated from this model look like? What are their sufficient statistics? Figure 2 is a plot of some characteristics of the

realizations of this model for different parameter values. Consider panel (a). It represents a plot of the parameter space of θ. The Bernoulli graphs correspond to the horizontal line at zero. Values in the bottom half-plane correspond to negatively dependent ties. Values in the top half-plane correspond to positively dependent ties. The gray-scale image is the probability of the empty graph graph for that parameter (i.e., $P_\theta(X = 0)$). Darker values correspond to higher probability. As expected, parameter values in the bottom-left quadrant place increasingly high probability on the empty graph.

Of what import is this? We rarely analyze network data when there are not ties (i.e., the graph is empty). Thus parameter values that correspond to distributions that say the empty graph is almost certain are not realistic. This is not to say that models should exclude completely the empty graph, only that models that place all their mass on the empty graph are not very interesting. Graph models in the bottom-left quadrant are examples of degenerate models for most empirical applications.

Consider panel (b). This is the equivalent plot for the probability of the complete graph. Parameter values in the top quadrants place high probability on the complete graph. These models are degenerate for most purposes. Panels (c)-(h) consider other forms of model degeneracy that may be unrealistic in many applications. Clearly the designation of model degeneracy depends on the use to which the model will be put, but the models in these panels are unlikely to be realistic for most applications.

Figure 3 is the cumulative plot for all the degeneracies in Figure 2. The models in the dark region place almost all their probability mass on a small number of unrealistic graph configurations. The vast majority of graphs, including almost all those likely to be of practical interest, have negligible probability of occurring. The light area corresponds to the set of models that are non-degenerate. Thus if we are searching for models that represent realistic behavior we must consider models in this small wedge of the parameter space. Thus even though any value of θ in $\mathbb{R}^2$ corresponds to a valid probability distribution, only this small subset of models will usually be "interesting" from a modeling viewpoint.

These figures illustrate that the *effective* natural parameter spaces of exponentially parametrized random graph models are limited. The shape of the non-degenerate region of the parameter space and especially the uneven shading of the bottom-right quadrant make direct interpretation of the region difficult.

For the $2-$star model the mean-value parametrization is

$$\mu_1 \equiv \mathbb{E}[E(x)] = \sum_{i<j} \mathbb{E}[x_{ij}] = N\mathbb{E}[x_{12}]$$

Here μ_1 is the expected number of edges, and μ_1/N is the probability that two actors are tied. Hence it is a direct and natural measure of network density on an interpretable scale. In addition it

is bounded between zero and N (or unity in the density form). The other parameter is

$$\mu_2 \equiv \mathbb{E}[S(x)] = \sum_{i<j<k} \mathbb{E}[x_{ij}x_{ik}] = 3\binom{g}{3}\mathbb{E}[x_{12}x_{13}]$$

Thus μ_2 is the expected number of $2-$stars, or $\mu_2/(3\binom{g}{3})$ is the probability that an actor is tied to randomly chosen other actors. Like μ_1, it is a direct and natural measure of network dependence on a interpretable scale. These advantages hold true for general models if the network statistics are directly interpretable.

Panel (a) of Figure 4 is a plot of the $\mu-$parameter space. The total shaded region (i.e., both red and green) is the parameter space and is identical to C, the convex hull of the sufficient statistics. The boundary of the space is marked by the sufficient statistics that form the convex hull of C. Note that the space is exactly the interior of the hull, that is, the continuous open bounded set.

As this space is just a mapping of the $\theta-$parameter space it must also contain the near degenerate graphs. However, they are not apparent. The red corresponds region in the plot corresponds to parameter values that are within 1 edge unit of the boundary. For example, values close to $(0, 0)$ correspond to probability distributions with close to zero edges and $2-$stars (i.e., empty graphs). Models with parameters close to $(21, 105)$ correspond to probability distributions with close to all edges and all $2-$stars (i.e., complete graphs). Models that hug the bottom of the hull have the minimum number of $2-$stars for the given number of edges. Finally models close to the top diagonal of the hull correspond to models with the maximum number of $2-$stars for the given number of edges. All of these models are near degenerate in the same sense as the previous section. However the geometry of the $\mu-$parameter space makes the identification of near degenerate models clearer. Different definitions of model degeneracy can be applied. The red band here represents models that expect to have less than one edge, at most one missing edge, or to be within one $2-$star of the extreme number for the given number of edges.

Equation (7) can be used to map non-degenerate models into the $\theta-$parameter space. Panel (b) of Figure 4 is a plot of the non-degenerate region of the $\theta-$parameter space mapped from panel (a). The interior of the "stealth-bomber"-like region corresponds to the parameter values that give non-degenerate models. Parameters outside this region fall in the border region of panel (a) and hence degenerate models by this definition. The crenelated nature of the region is one reason that analysis using the $\theta-$parameterization has been difficult, since only models in this region (or something similar in shape to it) are viable for modeling. The relative simplicity and interpretability of the $\mu-$parameter space to that in the $\theta-$parameter space is an important benefit. Of particular note is that small regions of the $\mu-$parameter space are mapped to large, unbounded regions of the $\theta-$parameter space. Panel (c) is a plot of the non-degenerate region of the natural parameter space with the location of the realizable MLEs superimposed. Most of the values for real data will exhibit positive dependence and occur in the top-left quadrant. In particular, many

will occur close to the narrow peninsula at the top (the "nose" of the stealth bomber). If a search routine does not take this geometry into account it will quickly yield issues of convergence.

Finally, panel (b) of Figure 4 can be overlayed with Figure 3. As can be seen the two regions are very similar. Thus this parametization provides the theoretical basis of the model degeneracy plots.

The behavior of the full conditional MCMC can be explained for the $2-$star model. For this model $M(\theta) = \max\{|\theta|, \theta_1 + 2(g-2)\theta_2\}$ and this corresponds to either the complete or empty graph. Hence if $\theta_1 >> 0$ the MCMC will have poor coverage of $\mathfrak{X}$. For example, consider the case where $g = 7$ and $\theta = (4.5, -18.4)$). For this model $\mu(\theta) \approx (3, 0)$ and $M(\theta) = 4.5$. An MCMC chain will approach $(3, 0) \in \text{rbd}(C)$ and stay there 98.9% of the time with 1.1% at $(2, 0) \in \text{rbd}(C)$; the sample has essentially two-states.

If $\theta_2 >> 0$ the MCMC will typically have poor coverage of $\mathfrak{X}$. Note that the magnitude of the second term in $M(\theta)$ increases linearly with the number of nodes in the graph. For example, consider in the case where $g = 7$ and $\theta = (-3.43, 0.683)$. For this model $\mu(\theta) \approx (9, 40)$ and $M(\theta) = \max\{3.43, 3.40)$. The distribution places 50% of its mass on graphs with 2 or fewer edges and 36% on graphs with at least 19 edges. It also mixes very slowly between these groups, as is shown, in Figure 5. This plots the traces and marginal densities of the samples of edges and $2-$stars from this model and based on 100,000 iterations of the full-conditional MCMC. The algorithm mixes very slowly between the set of nearly empty graphs and the nearly complete graphs. Alternative MCMC algorithms, such as a Swendsen-Wang variant (Swendsen and Wang 1987) would help in this situation, but not in general (e.g., the previous example).

The $2-$star model is unstable in the natural parametrization as small changes in the value of θ_2 when it is positive can greatly affect the model. For example, if $\theta_2 = 0.67$ in the previous model $\mu(\theta) \approx (4.4, 17.1)$ and the distribution places almost all its mass on the mode of empty graphs.

7. DISCUSSION

This paper addresses the question of what is a "good" model for a network and in doing so it helps to resolve the problem of lack of convergence when estimation or simulating using MCMC. These results also indicate that mean value and mixed parameterizations are a natural way to represent exponential parametrized random graph models. While they appear to suffer from more indirect specification of the likelihood, this is overshadowed by their interpretability, ease of estimation, and the stability and functionality of their parameter space.

One implication of these results is that the *effective* parameter space of exponentially parameterized random graph models is a small, bounded subset of the theoretical parameter space. Limiting inferences to this viable region in the parameter space can be accomplished in a Bayesian framework, via the use of parameter priors which concentrate most or all of the prior probability

mass on viable models. The Bayesian framework for inference promises to be very powerful for network modeling. In addition to reducing model degeneracy, it facilitates the propagation of parameter uncertainty into the final inference, and allows the incorporation of expert prior knowledge when it exists. While questions remain regarding the optimal selection of parameter priors, we expect that Bayesian inference for general random graph models will contribute greatly to the tools available to network researchers in the years ahead.

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. Our results, and the experience of others, suggest that commonly used random graph models are more global than local in structure, and that even those with nodal Markov dependence are global. Hence the current approach to pseudolikelihood estimation will have poor statistical properties. This same global nature of the existing models contributes to their model 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. Recent advances in model specification, cognizant of these issues, hold much promise (Nowicki and Snijders 2001; Pattison and Robins 2002; Hoff et al. 2002; Schweinberger and Snijders 2003; Snijders et al. 2004).

APPENDIX: TECHNICAL DETAILS

In this appendix we provide support for the results given in the paper. The model (1) is a finite statistical exponential family with respect to counting measure on $\mathfrak{X}$. We assume that the dimension of $\mathfrak{X}$ is q so that the family is minimal and the parameter space is $\{\theta : c(\theta, \mathfrak{X}) < \infty\} = \mathbb{R}^q$ so that it is also full. The results follow from the general theory of exponential families. Here we give a guide to those results.

Existence and uniqueness of the MPLE and MLE

The derivation of the result in Section 3.1 follows from the equivalence of the MPLE algorithm to a logistic regression on $\delta(x_{ij}^c$. This logistic model forms an exponential family. The result is then derivable from Barndorff-Nielsen (1978, Section 9.3). The equation of result 1 is derivable from the convex hull formed by $\delta(x_{ij}^c), \quad x \in \mathfrak{X}$. It can also be derived from Albert and Anderson (1984) and Santner and Duffy (1986). Our appeal to exponential family theory provides a direct alternative derivation to that given by those authors. The result in Section 3.2 follows from Barndorff-Nielsen (1978, Corollary 9.6).

Likelihood-based inference based on MCMC Algorithms

The sample $\{t(y^{(1)}), \ldots, t(y^{(M)})\}$ form a statistical exponential family with respect to counting measure. The result then follows from Barndorff-Nielsen (1978, Corollary 9.6).

What makes a useful model for a social network?

The first three statements are versions of Barndorff-Nielsen (1978, Section 9.8, (viii)). The last follows from the definition of the Kullback-Leibler discriminant information and the third result.

References

Albert, A. and Anderson, J. A. (1984), "On the existence of maximum likelihood estimates in logistic regression," *Biometrika*, 71, 1–10.

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

Baddeley, A. (2001), "Likelihoods and pseudolikelihoods for Markov spatial processes," in *State of the Art in Probability and Statistics: Festschrift for Willem R. van Zwet*, eds. Gunst, M., Klaassen, C., and Vaart, A., Institute of Mathematical Statistics, vol. 36 of *Lecture Notes–Monograph Series*, pp. 21–49.

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

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, 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 B*, 48, 259–302.

Borgatti, S. P., Everett, M. G., and Freeman, L. C. (1999), *Ucinet 5.0 for Windows*, Natick: Analytic Technologies.

Corander, J. Dahmstrom, K. and Dahmstrom, P. (1998), "Maximum Likelihood Estimation for Markov Graphs," Research report, Department of Statistics, University of Stockholm.

Crouch, B., Wasserman, S., 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 Sitges, Spain*.

Firth, D. (1993), "Bias Reduction in Maximum Likelihood Estimates," *Biometrika*, 80, 27–38.

Frank, O. (1971), *The Statistical Analysis of Networks*, London: Chapman and Hall.

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

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 B*, 54, 657–699.

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

Heinze, G. and Schemper, M. (2002), "A solution to the problem of separation in logistic regression," *Statistics in Medicine*, 21, 2409–2419.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97, 1090–1098.

Nowicki, K. and Snijders, T. A. B. (2001), "Estimation and Prediction for Stochastic Blockstructures," *Journal of the American Statistical Association*, 96, 1077–1087.

Pattison, P. and Robins, G. L. (2002), "Neighbourhood-based models for social networks," *Sociological Methodology*, 32, 301–337.

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

Santner, Thomas, J. and Duffy, D. E. (1986), "A Note on A. Albert and J. A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models," *Biometrika*, 73, 755–8.

Schweinberger, M. and Snijders, T. A. B. (2003), "Settings in Social Networks: A Measurement Model," *Sociological Methodology*, 33, 307–341.

Snijders, T. A. B. (2002), "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 3.

Snijders, T. A. B., Pattison, P., Robins, G. L., and Handcock, M. S. (2004), "New specifications for exponential random graph models," Working paper, ICS, Department of Sociology, University of Groningen.

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.

Strauss, D. J. (1975), "A model for clustering," *Biometrika*, 62, 467–475.

Swendsen, R. H. and Wang, J.-S. (1987), "Non-universal critical dynamics in Monte Carlo simulation," *Physical Review Letters*, 88, 58–86.

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.

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

Wasserman, S. S. and Robins, G. L. (2004), "An Introduction to Random Graphs, Dependence Graphs, and p^*," in *Models and Methods in Social Network Analysis*, eds. P. Carrington, J. S. and Wasserman, S. S., Cambridge: Cambridge University Press, p. in press.

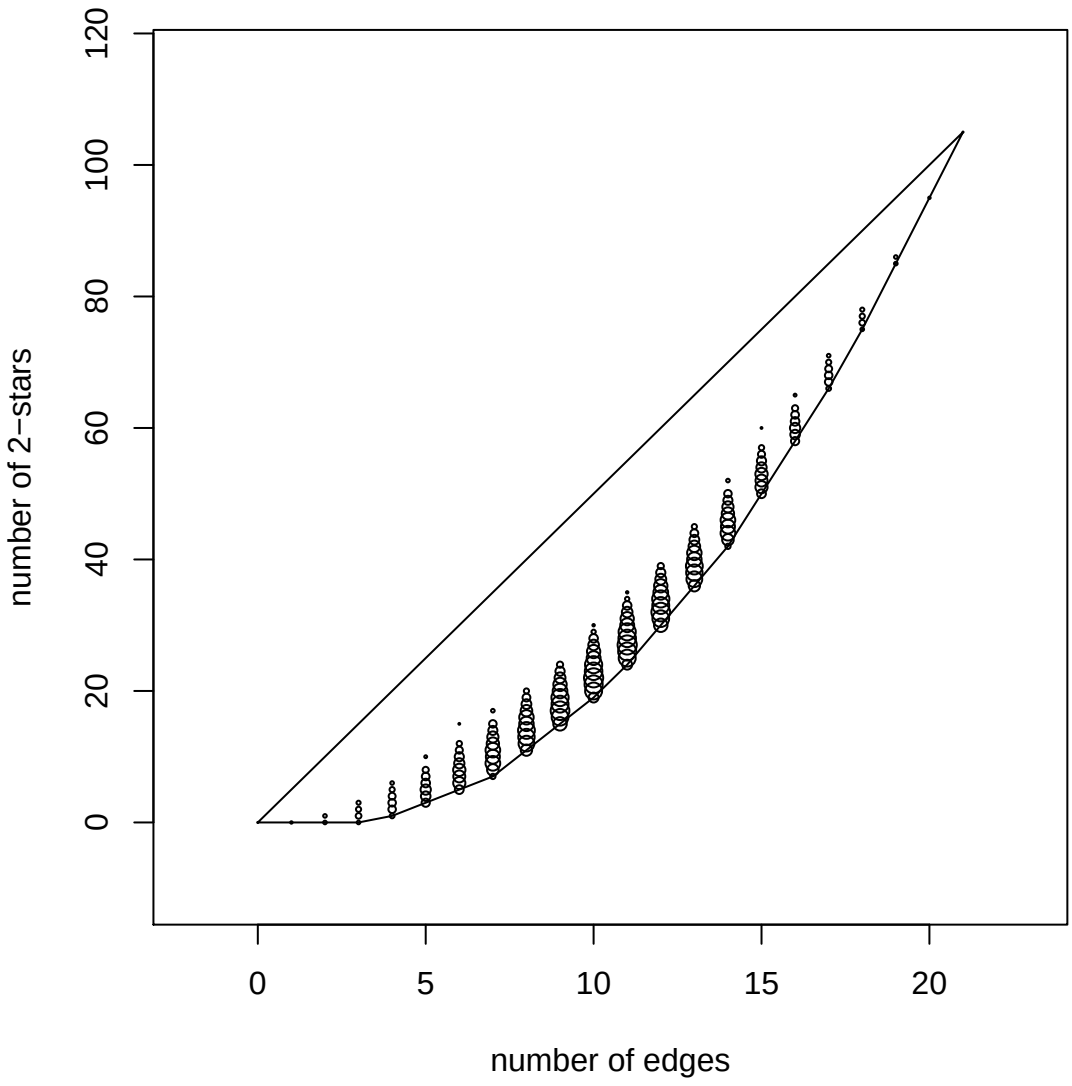


Figure 1: Enumeration of sufficient statistics for graphs with 7 nodes. The circles are centered on the possible values and the area of the circle is proportional to the number of graphs with that value of the sufficient statistic. There are a total of 2,097,152 graphs.

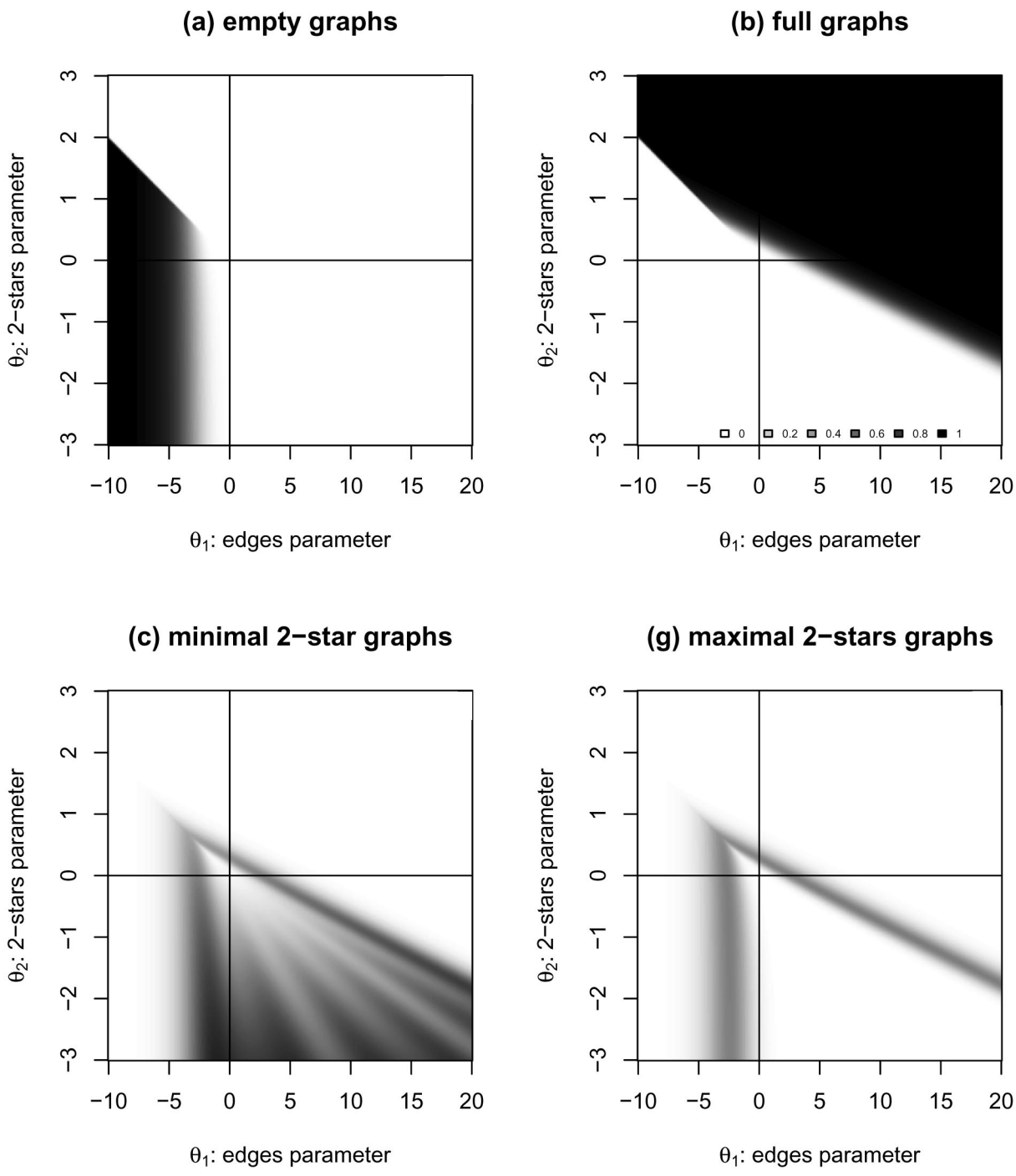
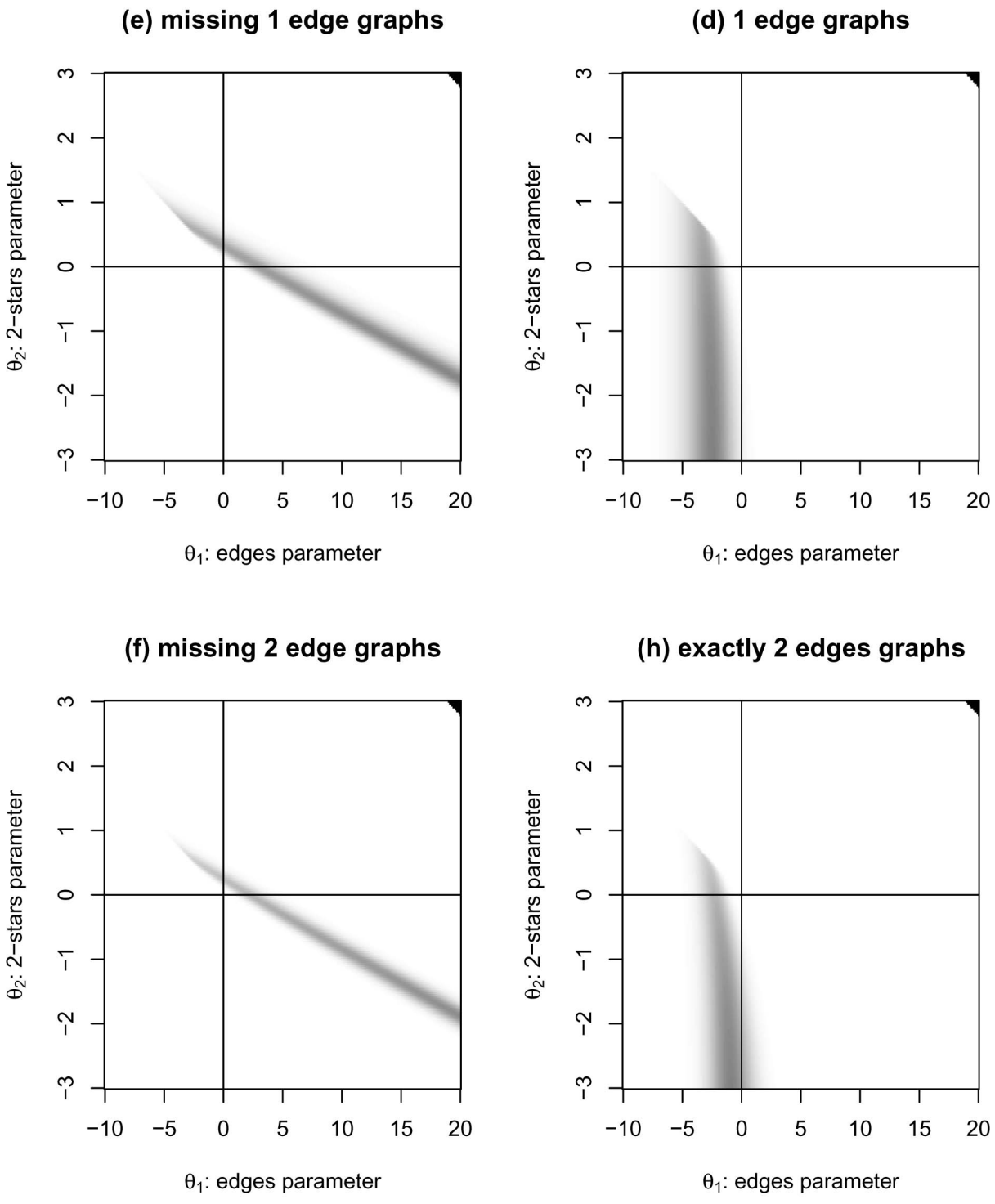


Figure 2: Degeneracy probability plots for graphs with 7 actors.


(e) missing 1 edge graphs
(d) 1 edge graphs
(f) missing 2 edge graphs
(h) exactly 2 edges graphs
θ2: 2−stars parameter
θ1: edges parameter
−10 −5 0 5 10 15 20
−3 −2 −1 0 1 2 3

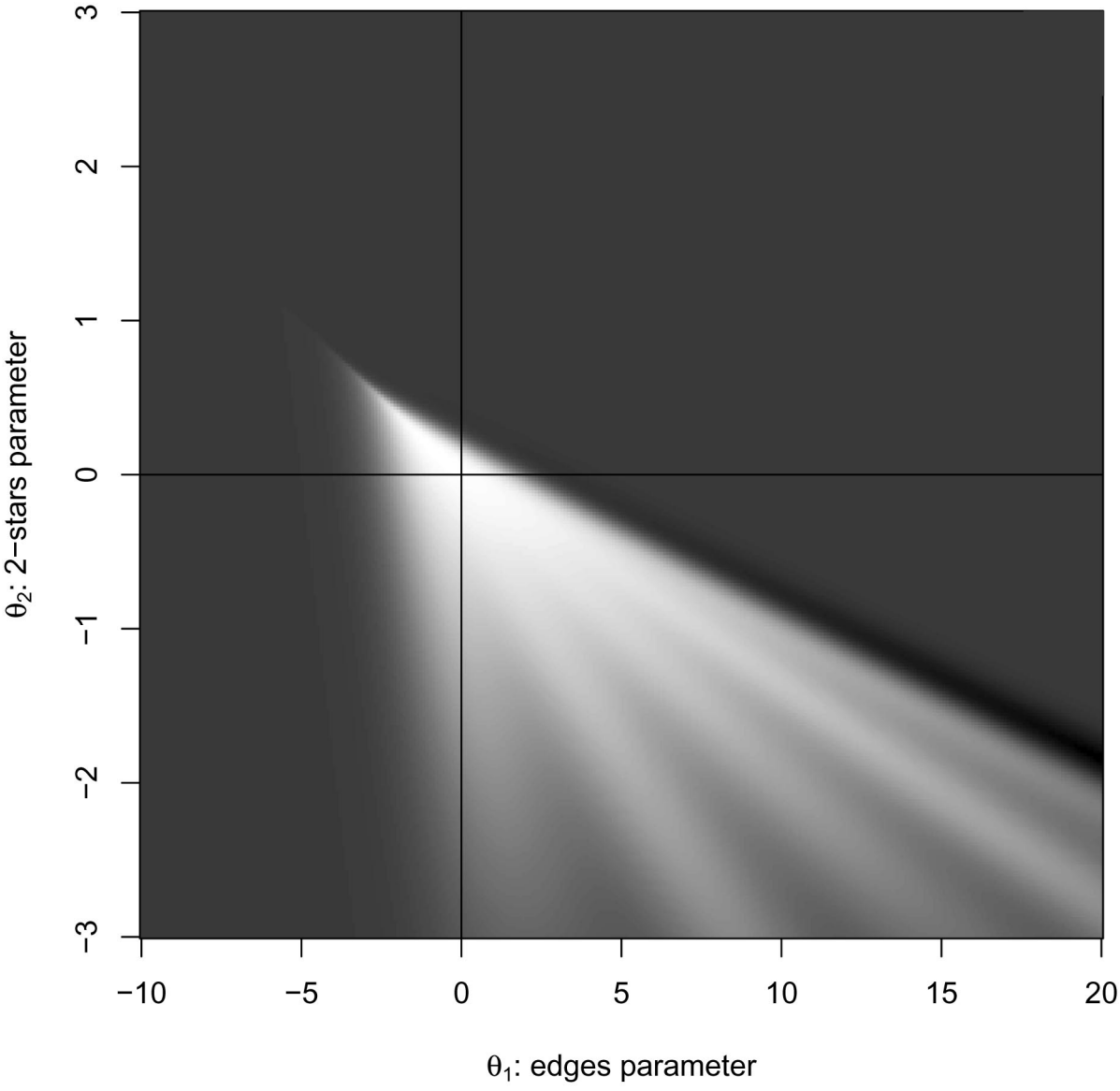


Figure 3: Cumulative Degeneracy Probabilities for graphs with 7 actors.

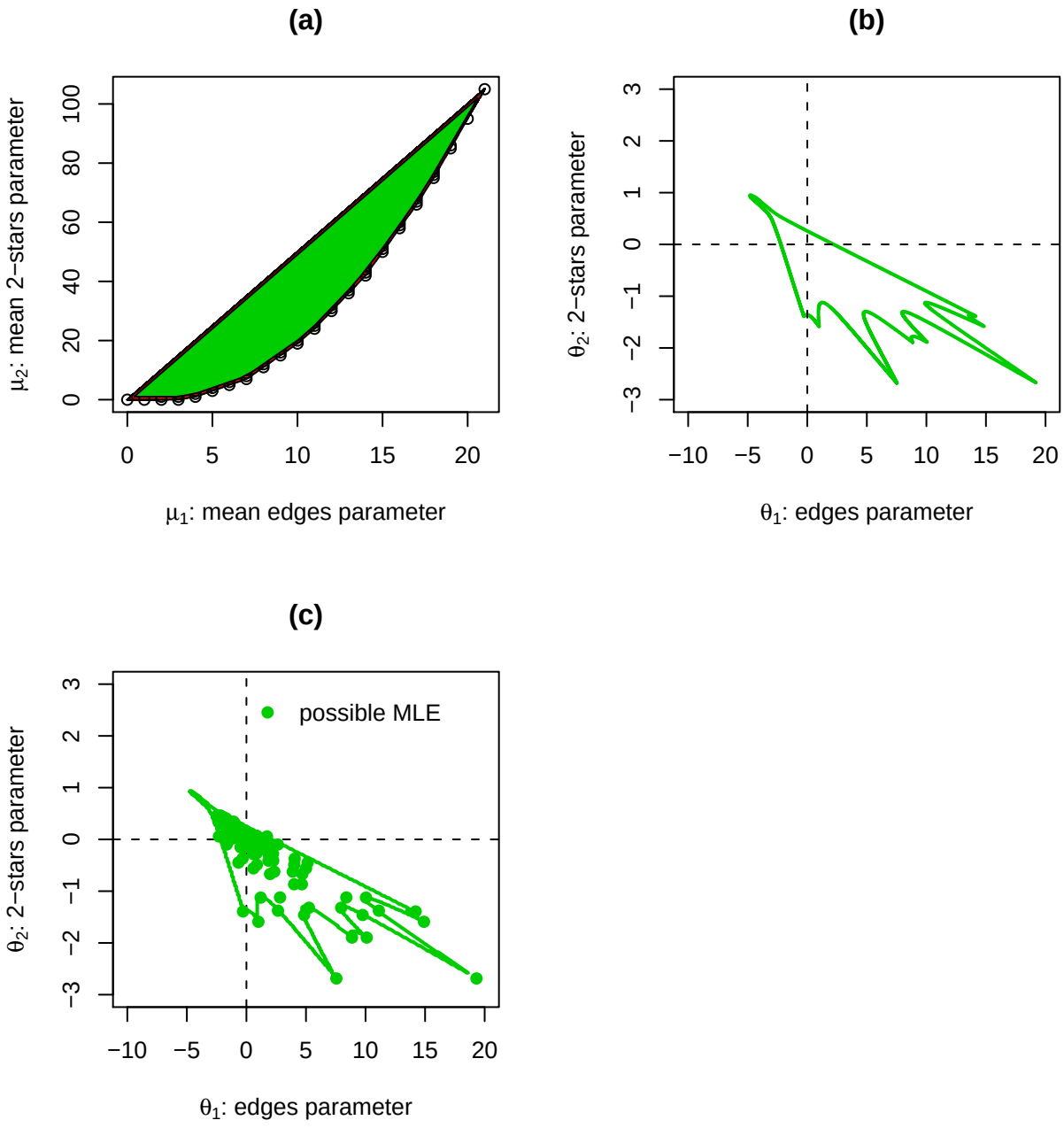


Figure 4: Regions of the mean value and natural parameter spaces.

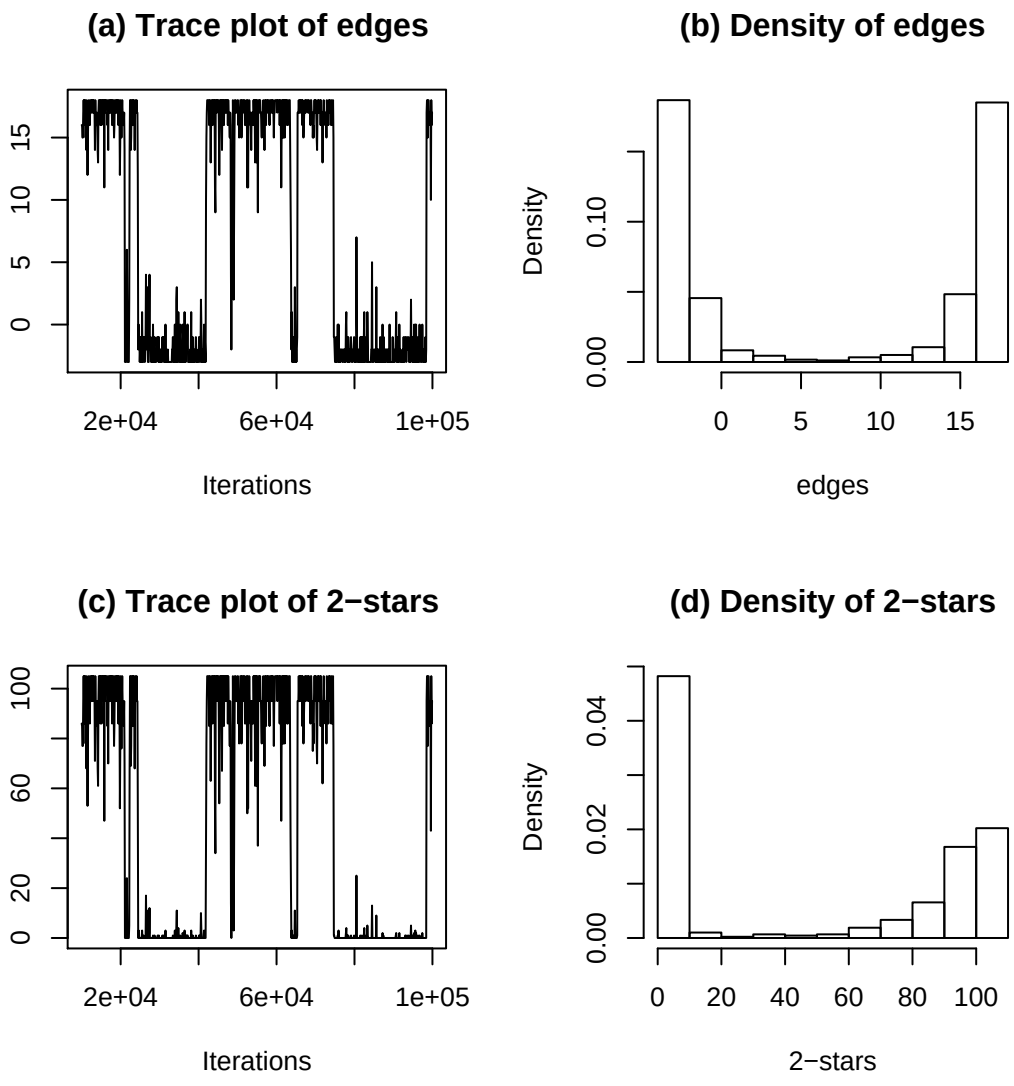


Figure 5: MCMC diagnostic plots for the model with $\theta = (-3.43, 0.683)$.