Title
A curved exponential family model for complex networks

Journal
Computational and Mathematical Organization Theory, 15(4)

ISSN
1381-298X

Authors
Handcock, Mark S
Morris, Martina

Publication Date
2009-12-01

DOI
10.1007/s10588-008-9055-x

Peer reviewed

Comput Math Organ Theory (2009) 15: 294–302
DOI 10.1007/s10588-008-9055-x

A curved exponential family model for complex networks

Mark S. Handcock · Martina Morris

Published online: 1 November 2008

Abstract Networks are being increasingly used to represent relational data. As the patterns of relations tends to be complex, many probabilistic models have been proposed to capture the structural properties of the process that generated the networks. Two features of network phenomena not captured by the simplest models is the variation in the number of relations individual entities have and the clustering of their relations. In this paper we present a statistical model within the curved exponential family class that can represent both arbitrary degree distributions and an average clustering coefficient. We present two tunable parameterizations of the model and give their interpretation. We also present a Markov Chain Monte Carlo (MCMC) algorithm that can be used to generate networks from this model.

Keywords Random graph models · Markov chain Monte Carlo · Statistical exponential families

1 Introduction

The characterization of structural properties of network data is of great interest. Network data arise in many fields of study and the traditional methods of statistical summarization struggle to capture the observed complexities. The phenomena we study are mainly from the social sciences and the forms of relations are social. In these social networks the nodes represent individual people and the relations represent some form of social contact or partnership. Here we assume that the network is a realization of a stochastic process and am primarily interested in those characterized by random mixing between individuals conditional on the individual activity levels (i.e., the

M.S. Handcock (✉) · M. Morris
Department of Statistics, University of Washington, Seattle, USA
e-mail: handcock@stat.washington.edu

nodal degrees) and clustering (Newman 2002; Dezső and Barabási 2002). One popular class of models are those that exhibit power-law behavior, often loosely referred to as "scale-free" distributions. We also consider models for the network degree distributions in which the variance can greatly exceed the mean. Our model generalizes models based solely on the degree distribution.

In Sect. 2 we develop the general form of the structural model and specific models for the degree distribution. In Sect. 3 we give a simple algorithm for the generation of random networks from the model. In Sect. 4, we discuss generalizations of the model for more complex structures.

2 Models for social networks

We first review the (linear) exponential family class of models, then we introduce the sub-class that is the focus of this paper. We then review the specification of the model that corresponds to the degree distribution of the network.

2.1 Exponential family models

Let the random matrix Y represent the adjacency matrix of an unvalued network on n individuals. We assume that the diagonal elements of Y are 0—that self-partnerships are disallowed. Suppose that $\mathbb{Y}$ denotes the set of all possible networks on the given n individuals. The multivariate distribution of Y can be parameterized in the form:

$$P_{\eta,\mathbb{Y}}(Y = y) = \frac{\exp[\eta \cdot Z(y)]}{c(\eta, \mathbb{Y})} \quad y \in \mathbb{Y} \tag{1}$$

where $\eta \in \Upsilon \subseteq \mathbb{R}^q$ is the model parameter and $Z{:}\mathbb{Y} \to \mathbb{R}^q$ are statistics based on the adjacency matrix (Frank and Strauss 1986; Handcock 2002). 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., centrality, mutuality and betweenness) of primary substantive interest to the researcher. In many situations the researcher has specified a set of statistics based on substantive theoretical considerations. The above model then has the property of maximizing the entropy within the family of all distributions with given expectation of $Z(Y)$ (Barndorff-Nielsen 1978). Paired with the flexibility of the choice of Z this property does provide some justification for the model (1) that will vary from application to application.

The denominator $c(\eta, \mathbb{Y})$ is the normalizing function that ensures the distribution sums to one: $c(\eta, \mathbb{Y}) = \sum_{y \in \mathbb{Y}} \exp[\eta \cdot Z(y)]$. This factor varies with both η and the support $\mathbb{Y}$ and is the primary barrier to simulation and inference under this modeling scheme.

The most commonly used class of random network models exhibit 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 networks have the same probability under the model. It is

shown in Frank and Strauss (1986) that the class of homogeneous Markov undirected networks is exactly those having the *degree parameterization*:

$$d_k(y) = \text{the proportion of nodes with degree exactly } k \qquad k = 0, \ldots, n-1$$

$$T_\Delta(y) = \frac{1}{6} \sum_{i,j,k} y_{ij} y_{jk} y_{kl}$$

where $d_k(y)$ counts the proportion of individuals with degree k and $T_\Delta(y)$ is a count of the complete triads. Throughout we consider undirected networks, although the situation for directed networks is very similar. This model can be reexpressed in the notation of model (1) by setting $Z_k(y) = d_k(y),\ k = 1, \ldots, n-1,\ Z_n = T_\Delta(y),\ q = n,$ $\eta \in \Upsilon = \mathbb{R}^n$. This parameterization has the advantage that it is directly interpretable in terms of concurrency of partnerships (i.e. $d_m(y)$ for $m > 0$ is the proportion of individuals with exactly m concurrent partners).

A popular variant of the statistic $T_\Delta(y)$ is the clustering coefficient defined as

$$C(y) = \frac{3T_\Delta(y)}{S_2(y)}$$

where $S_2(y)$ is the number of connected triples of nodes (i.e., 2-stars, Frank and Strauss 1986). This describes the proportion of complete triads in the networks out of a total number of possible triads.

2.2 A model for the degree distribution and clustering

In the remainder of this paper we focus on the sub-class of the exponential family given by the model:

$$\log[P_\theta(Y = y)] = \eta(\phi) \cdot d(y) + \nu C(y) - \log c(\phi, \nu, \mathbb{Y}) \tag{2}$$

where $y \in \mathbb{Y}, \theta = (\phi, \nu), \Theta \subset \mathbb{R}^n,\ d(y) = \{d_1(y), \ldots, d_{n-1}(y)\}$. The parameters ϕ and ν represent the network degree distribution and clustering, respectively. Specifically, the ratio of the probability of a given network to a network with the same degree distribution and correlation coefficient 1% less is $0.01 \times \exp(\nu)$. Alternatively, consider the conditional probability of a partnership existing given the rest of the network. If the formation of the partnership increases the correlation coefficient by $\alpha\%$ (relative to the same network without the partnership) then the log-odds of the partnership existing is $\alpha\nu\%$. The degree distribution parameters have similar interpretations: $\eta_k(\phi)$ is the ratio of the log-probability of a given network to a network with the same clustering coefficient and one less node of degree k and one more isolate. An important property of the model is the variational independence of the parameters (Barndorff-Nielsen 1978).

This model is a curved exponential family if Θ is a smooth curve in $\Upsilon = \mathbb{R}^n$ (Hunter and Handcock 2006; Handcock 2003a). Any degree distribution can be specified by $n-1$ or less independent parameters. Typically the number of parameters is small. As we shall see, this is true for the models considered below.

If $\nu = 0$ the model corresponds to random networks with arbitrary degree distributions, as considered by many researchers (Newman et al. 2001). If $\eta_k(\phi) = \phi k$, $k = 1, \ldots, n-1$ the value of ϕ is interpretable as the log-probability of a given network to a network with one less partnership and the same clustering coefficient (Hunter and Handcock 2006). If both $\nu = 0$ and $\eta_k(\phi) = \phi k$, $k = 1, \ldots, n-1$ it is the classical random network model of Rényi and Erdös (Bollobas 1985).

The model (1) has a generative interpretation, which we illustrate with model (2). Consider a dynamic process for the network $\{Y(t)\colon t \geq 0\}$ developing according to the local rules

$$\text{logit}[P(Y_{ij}(t) = 1 | Y_{ij}(t^-) = y_{ij})] = \eta(\phi) \cdot [d(y_{ij}^+) - d(y_{ij}^-)] + \nu[C(y_{ij}^+) - C(y_{ij}^-)]$$

where y_{ij}^+ is the network with a partnership between i and j and the rest of the network equal to y_{ij}. y_{ij}^- is similar with no partnership between i and j. Based on the theory of continuous-time Markov Chains, the equilibrium distribution is model (2). Ties are formed (or broken) based on their propensity to change the network characteristics. This also provides another interpretation of the parameters ϕ and ν and their joint effects.

An alternative parameterization that is usually more interpretable is: (ϕ, ρ) where the mapping is:

$$\rho = \mathbb{E}_{\phi,\rho}[C(Y)] = \sum_{y \in \mathbb{Y}} C(y) \exp[\eta(\phi) \cdot d(y) + \nu C(y)] \geq 0 \tag{3}$$

Thus ρ is the mean clustering coefficient over networks in $\mathbb{Y}$. Thus models with higher ρ have higher clustering coefficients on average. Note that models with $\rho = 0$ will not have any complete triads. The range of ρ is a subset of $[0, 1]$ and depends on the other parameters and $\mathbb{Y}$.

The two parameterizations represent the same model class (Handcock 2003a). Translating between equivalent parameters is achieved using the MCMC algorithm given in Sect. 3 (Handcock 2003a; Hunter and Handcock 2006).

2.3 Models for degree distributions

Let $P_\theta(K = k)$ be the probability mass function of K, the number of partnerships that a randomly chosen node in the network has. Based on the model (2)

$$P_\theta(K = k) = \mathbb{E}_\theta[d_k(Y)] \quad k = 0, \ldots, n-1$$

Clearly for a given network of size n nodes, the distribution of K has finite range with upper bound $n-1$. In some cases this distribution is approximated by an idealized distribution with infinite range. Let K^* be the degree of a node in a (possibly hypothetical) infinite population of nodes. Then K can be thought of as the degree of the node restricted to nodes in the network. In cases where this conceptualization is used we will consider the case

$$P_\theta(K = k) = P(K^* = k | K^* < n) \quad k = 0, \ldots, n-1$$

While the model (2) has arbitrary degree distribution, of particular interest are the various "scale-free," preferential attachment and power-law models popular in the physics literature (see, e.g., Newman 2003). These models assume that all networks with the same degree distribution are equally likely. We say $P(K^* = k)$ has *power-law behavior* with scaling exponent $\phi > 1$ if there exist constants c_1, c_2, and M such that $0 < c_1 \leq P(K^* = k)k^{\phi} \leq c_2 < \infty$ for $k > M$.

We focus on a stochastic mechanisms for the formation of the social networks that is a variation on a preferential attachment process, such as those advocated by several recent authors (Barabási and Albert 1999; Pastor-Satorras and Vespignani 2001). The limiting distributions of this mechanism can be characterized by long tails.

2.4 Simple preferential attachment models

A mechanism that has been suggested for the formation of power-law social networks is preferential attachment (Albert and Barabási 2000; Liljeros et al. 2001; Dezső and Barabási 2002). This and related stochastic processes have a long history in applied statistics (Simon 1955; Kendall 1961; Irwin 1963). Consider the formation process of partnerships within an infinite population of people. The partnerships form so that: (1) there is a constant probability p that the $r + 1$st partnership in the population will be initiated from a randomly chosen person to a person with zero partnerships, and (2) otherwise the probability that the $r + 1$st partnership will be to a person with exactly k partnerships is proportional to $kf(k|r)$, where $f(k|r)$ is the frequency of people with exactly k partnerships out of the r total partnerships in the population. The limiting distribution of the (marginal) number of partnerships from this process is known as the Waring distribution (Irwin 1963). The Yule distribution discussed by Simon (1955) and used by Jones and Handcock (2003) to model degree distributions is a special case of the Waring distribution with $p = (\phi_2 - 2)/(\phi_2 - 1)$.

The probability mass function (PMF) of the Waring distribution (Johnson et al. 1992) is:

$$P(K^* = k) = \frac{(\phi_2 - 1)\Gamma(\phi_2 + \phi_1)}{\Gamma(\phi_1 + 1)} \cdot \frac{\Gamma(k + \phi_1)}{\Gamma(k + \phi_1 + \phi_2)}$$
$$\phi_1 > -1, \phi_2 > 2 \tag{4}$$

where $\Gamma(\cdot)$ is the Gamma function and the mixing parameter ϕ_1 is related to p via:

$$p = \frac{\phi_2 - 2}{\phi_2 + \phi_1 - 1} \tag{5}$$

The Waring distribution has power-law behavior with scaling exponent ϕ_2. The mean and variance of the Waring distribution are:

$$\mathbb{E}(K^*) = \frac{1}{p}, \qquad \mathbb{V}(K^*) = \frac{(1 - p)(\phi_2 - 1)}{p^2(\phi_2 - 3)}, \quad \phi_2 > 3$$

Thus, the expected value of the Waring distribution is simply the inverse of the probability of forming a partnership to an individual lacking existing partnerships.

3 Generating random networks with specified structure

In this section we review Markov Chain Monte Carlo (MCMC) algorithms for network generation and present a two-stage variant algorithm that may be more efficient.

MCMC algorithms for generating from the model (1) have a long history and been well studied (see Geyer and Thompson 1992 for a review). The basic idea is to generate a Markov chain whose stationary distribution is given by (1). The simplest Markov chain proceeds by choosing (by some method, either stochastic or deterministic) a dyad (i, j) and then deciding whether to set $Y_{ij} = 1$ or $Y_{ij} = 0$ at the next step of the chain. One way to do this is using Gibbs sampling, whereby the new value of Y_{ij} is sampled from the conditional distribution of Y_{ij} conditional on the rest of the network. Denote "the rest of the network" by Y^c_{ij}. Then $Y_{ij}|Y^c_{ij} = y^c_{ij}$ has a Bernoulli distribution, with odds given by

$$\frac{P(Y_{ij} = 1|Y^c_{ij} = y^c_{ij})}{P(Y_{ij} = 0|Y^c_{ij} = y^c_{ij})} = \exp\{\eta \cdot \Delta(Z(y))_{ij}\}$$

where $\Delta(Z(y))_{ij}$ denotes the difference between $Z(y)$ when y_{ij} is set to 1 and $Z(y)$ when y_{ij} is set to 0. A simple variant to the Gibbs sampler (which is an instance of a Metropolis-Hastings algorithm) is a pure Metropolis algorithm in which the proposal is always to change the value of y_{ij}. This proposal is accepted with probability $\min\{1, \pi\}$, where

$$\begin{aligned} \pi &= \frac{P(Y_{ij} = 1 - y_{ij}|Y^c_{ij} = y^c_{ij})}{P(Y_{ij} = y_{ij}|Y^c_{ij} = y^c_{ij})} \\ &= \begin{cases} \exp\{\eta \cdot \Delta(Z(y))_{ij}\} & \text{if } y_{ij} = 0 \\ \exp\{-\eta \cdot \Delta(Z(y))_{ij}\} & \text{if } y_{ij} = 1 \end{cases} \end{aligned} \tag{6}$$

The vector $\Delta(Z(y))_{ij}$ used by these MCMC schemes is often much easier to calculate directly than as the difference of two separate values of $Z(y)$. For instance, if one of the components of the $Z(y)$ vector is the total number of partnerships in the network, then the corresponding component of $\Delta(Z(y))_{ij}$ is always equal to 1.

The Metropolis scheme is usually preferred over the Gibbs scheme because it results in a greater probability of changing the value of y_{ij}, a property thought to produce better-mixing chains. However, it is well known that these simple MCMC schemes often fail for various reasons to produce well-mixed chains (Snijders 2002; Handcock 2000; Snijders et al. 2006). More sophisticated MCMC schemes have been developed and are a topic of ongoing research (Hunter and Handcock 2006).

A variant of this algorithm proceeds in two steps:

1. Generate $d_k \overset{\text{i.i.d.}}{\sim} P_\theta(K = k),\ k = 0, 1, \ldots, n-1$.
2. Generate a random network conditional on this degree distribution:

$$P_\nu(Y = y|d_k(Y) = d_k) = \frac{\exp[\nu C(y)]}{c(\nu, d_k, \mathbb{Y})} \quad y \in \mathbb{Y}(d_k)$$

where $\mathbb{Y}(d_k) = \{y \in \mathbb{Y}: d_k(y) = d_k\}$.

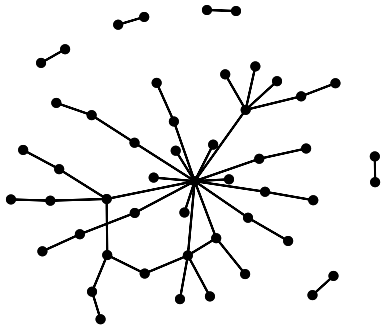

Fig. 1 An example network generated from model (2) with $n = 50$ and degree distribution draw from the Yule model (4) with scaling exponent $\phi_2 = 3$. The random network is drawn from the model with mean clustering coefficient $\rho = 3\%$. The network has clustering coefficient $C(y) = 2\%$

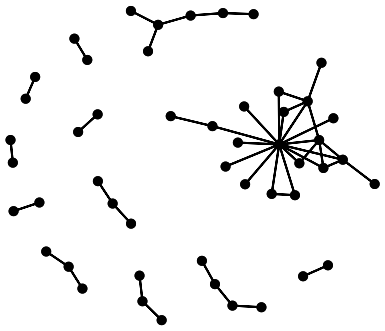

Fig. 2 An example network generated from model (2) with $n = 50$ and degree distribution draw from the Yule model (4) with scaling exponent $\phi_2 = 3$. The random network is drawn from the model with mean clustering coefficient $\rho = 15\%$. The network has clustering coefficient $C(y) = 18\%$

The first generates individual degrees from an arbitrary distribution, and the second generates networks conditioned on those degrees. Note that the structure of the exponential family in (1) ensures that the samples are from the correct distribution (Barndorff-Nielsen 1978). The first step can be simulated easily as we know $P_\theta(K = k)$. Note that not all degree sequences will be consistent with a network of size n. For example, sequences with an odd total number of partnerships are not realizable. However we can construct a compatible sequence $\{d_k\}_{k=0}^{n-1}$ via a simple rejection algorithm. The second step is also straightforward: we can conditionally simulate values using a MCMC holding the degree distribution fixed by using a Metropolis proposal consistent with this restriction. It is convenient for this algorithm to have a starting network with the given degree distribution. This network is easy to construct by a finite algorithm (as it need not be a draw from a random distribution) or using sequential importance sampling. An important property of this second step is the independence of the distribution from ϕ. It is a simple parameter distribution depending only on ν (Barndorff-Nielsen 1978).

As an application of this algorithm, consider a network model for $n = 50$ nodes. We choose a degree distribution which is Yule with scaling exponent $\phi_2 = 3$. This corresponds to a "scale-free" degree model. If $\nu = 0$ the network is random with the given degree distribution. This corresponds to a mean clustering coefficient $\rho = 3\%$. A realization of this model is given in Fig. 1. The clustering coefficient for this network is 2%. Figure 2 is a realization from the model with mean clustering coefficient $\rho = 15\%$ (corresponding to a clustering parameter of $\nu = 0.46$). The centralization of the clustering is apparent relative to the network in Fig. 1.

An R package called statnet has been written to implement the procedures in this paper (Handcock et al. 2003). statnet is an open-source software suite for network modeling and is written in the R statistical language (R Development Core Team 2008).

4 Discussion

Models that represent the degree distribution of a network often fail to adequately represent the level of clustering in network phenomena. We have presented a simple stochastic model for random networks that has arbitrary degree distribution and average clustering coefficient. The clustering component of the model is directly interpretable via the clustering coefficient of the realizations from the model. The model places positive probability over the set of possible networks. Conditional on the degree sequence, the clustering coefficient covers the full range of values possible. The distribution over this range is tuned as a monotone function of the clustering parameter.

We note that the model form (1) is very general, and can incorporate general social structure (Frank and Strauss 1986; Strauss and Ikeda 1990; Handcock 2003a; Hunter and Handcock 2006). For example, in disease epidemiology, the two-sex random network epidemic model is a commonly used to represent the contact structure of pathogens transmitted by intimate contact. This model is the model (2) with $\rho = 0$ and $\mathbb{Y}$ is restricted to heterosexual networks. However, this model contains a major weakness which ultimately limits its utility. Specifically, it assumes random mixing conditional on degree. The model (2) is a simple extension of that allows tunable correlation coefficient. More generally, (1) can be used to include nodal attributes and other structural characteristics. Such models have proven to be valuable in epidemiology (Morris 2003; Handcock 2003b).

Acknowledgements This research is supported by Grant DA012831 from NIDA and Grant HD041877 from NICHD. The authors are grateful to David Hunter and David Banks for discussions and helpful suggestions. Parts of this paper were in response to ideas presented at the Workshop about power-law models. An extended version of this paper was subsequently presented at the "Statistical Network Analysis: Models, Issues, and New Directions" Workshop at the 23rd International Conference on Machine Learning (ICML 2006), June 29, 2006, Pittsburgh PA, USA.

References

Albert R, Barabási AL (2000) Topology of evolving networks: Local events and universality. Phys Rev Lett 85(24):5234–5237

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Barndorff-Nielsen OE (1978) Information and exponential families in statistical theory. Wiley, New York

Bollobas B (1985) Random graphs. Academic Press, New York

Borgatti SP, Everett MG, Freeman LC (1999) Ucinet 6.0 for Windows. Analytic Technologies, Natick

Dezső Z, Barabási AL (2002) Halting viruses in scale-free networks. Phys Rev E 65:055103

Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81(395):832–842

Geyer CJ, Thompson EA (1992) Constrained Monte Carlo maximum likelihood calculations (with discussion). J R Stat Soc Ser B 54:657–699

Handcock MS (2000) Progress in statistical modeling of drug user and sexual networks. Manuscript, Center for Statistics and the Social Sciences, University of Washington
Handcock MS (2002) Degeneracy and inference for social network models. Paper presented at the Sunbelt XXII international social network conference, New Orleans, LA
Handcock MS (2003a) Assessing degeneracy in statistical models of social networks. Working paper #39, Center for Statistics and the Social Sciences, University of Washington
Handcock MS (2003b) Statistical models for social networks: Inference and degeneracy. In: Breiger R, Carley K, Pattison P (eds) Dynamic social network modeling and analysis. Committee on human factors, board on behavioral, cognitive, and sensory sciences. National Academy Press, Washington, pp 229–240
Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2003) `statnet`: Software tools for the statistical modeling of network data. Statnet Project, Seattle, WA. Version 2.0. Project home page at http://statnet.org, http://CRAN.R-project.org/package=statnet
Hunter DR, Handcock MS (2006) Inference in curved exponential family models for networks. J Comput Graph Stat 15:565–583
Irwin J (1963) The place of mathematics in medical and biological statistics. J R Stat Soc Ser A (General) 126(1):1–45
Johnson N, Kotz S, Kemp A (1992) Univariate discrete distributions. Wiley series in probability and mathematical statistics, 2nd edn. Wiley, New York
Jones J, Handcock MS (2003) An assessment of preferential attachment as a mechanism for human sexual network formation. Proc R Soc Lond B 270:1123–1128
Kendall M (1961) Natural law in the social sciences: Presidential address. J R Stat Soc Ser A—Stat Soc 124(1):1–16
Liljeros F, Edling CR, Amaral LAN, Stanley HE, Åberg Y (2001) The web of human sexual contacts. Nature 411(6840):907–908
Morris M (2003) Local rules and global properties: Modeling the emergence of network structure. In: Breiger R, Carley K, Pattison P (eds) Dynamic social network modeling and analysis. Committee on human factors, board on behavioral, cognitive, and sensory sciences. National Academy Press, Washington, pp 174–186
Newman MEJ (2002) Spread of epidemic disease on networks. Phys Rev E 66(1):016128
Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256
Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Phys Rev E 64:026118
Pastor-Satorras R, Vespignani A (2001) Epidemic dynamics and endemic states in complex networks. Phys Rev E 63(6):066117
R Development Core Team R: (2008) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 2.8.0. ISBN 3-900051-07-0. http://www.R-project.org
Simon H (1955) On a class of skew distribution functions. Biometrika 42(3/4):435–440
Snijders TAB (2002) Markov chain Monte Carlo estimation of exponential random graph models. J Soc Struct 3(2)
Snijders TAB, Pattison P, Robins GL, Handcock MS (2006) New specifications for exponential random graph models. Sociol Methodol 36:99–153
Strauss D, Ikeda M (1990) Pseudolikelihood estimation for social networks. J Am Stat Assoc 85:204–212
Wasserman SS, Faust K (1994) Social network analysis: Methods and applications. Cambridge University Press, Cambridge

Mark S. Handcock is a Professor of Statistics, Department of Statistics, University of Washington, Seattle. His work focuses on the development of statistical models for the analysis of social network data, spatial processes and demography. He received his B.Sc. from the University of Western Australia and his Ph.D. from the University of Chicago. Descriptions of his work are available at http://www.stat.washington.edu/handcock.

Martina Morris is a Professor of Sociology and Statistics at the University of Washington. Her work focuses on the development of statistical models for the analysis of social network data, the demographic epidemiology of HIV, trends in earnings inequality, and innovative statistical methodology for demographic research. She received his B.Sc. from the Reed College and her Ph.D. from the University of Chicago. Descriptions of her work are available at http://faculty.washington.edu/morrism/?UID=morris.