Open access • Book Chapter • DOI:10.1007/978-3-540-73133-7_8

A simple model for complex networks with arbitrary degree distribution and clustering — Source link

Mark S. Handcock, Martina Morris

Institutions: University of Washington

Published on: 29 Jun 2006 - International Conference on Machine Learning

Topics: Degree distribution, Hierarchical network model, Complex network, Statistical model and Cluster analysis

A Simple Model for Complex Networks with Arbitrary Degree Distribution and Clustering

Mark S. Handcock and Martina Morris*

University of Washington, Seattle WA 98195-4322, USA
handcock@stat.washington.edu
http://www.stat.washington.edu/~handcock

Abstract. We present a stochastic model for networks with arbitrary degree distributions and average clustering coefficient. Many descriptions of networks are based solely on their computed degree distribution and clustering coefficient. We propose a statistical model based on these characterizations. This model generalizes models based solely on the degree distribution and is within the curved exponential family class. We present alternative parameterizations of the model. Each parameterization of the model is interpretable and tunable. We present a simple Markov Chain Monte Carlo (MCMC) algorithm to generate networks with the specified characteristics. We provide an algorithm based on MCMC to infer the network properties from network data and develop statistical inference for the model. The model is generalizable to include mixing based on attributes and other complex social structure. An application is made to modeling a protein to protein interaction network.

1 Introduction

In this paper we consider models for relational data, and specifically networks. We have in mind social networks where the nodes represent individuals and the edges represent some form of social contact or partnership. However, the formulation is general and can be used to represent other forms of networks. We assume that the network is a realization of a stochastic process characterized by random mixing between individuals conditional on the individual activity levels (i.e., the nodal degrees) and clustering [1, 2]. One popular class 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.

In Section 2 we develop the general form of the model and models for the degree distribution. In Section 3 we give an simple algorithm for the generation of random networks from the model. In Section 4 we provide an algorithm for approximating the likelihood function for the model as a basis for inference. In Section 5 we apply the model to a protein-protein interaction network. Finally, in Section 6, we discuss generalizations of the model for more complex structures.

* We gratefully acknowledge the critical feedback we have received from David Hunter, Steve Goodreau and Carter T. Butts. This research supported by Grant DA012831 from NIDA and Grant HD041877 from NICHD.

E.M. Airoldi et al. (Eds.): ICML 2006 Ws, LNCS 4503, pp. 103–114, 2007.

2 Models for Social Networks

2.1 Exponential Family Models

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

$$P_{\eta,\mathcal{X}}(X = x) = \frac{\exp\left[\eta \cdot Z(x)\right]}{c(\eta, \mathcal{X})} \qquad x \in \mathcal{X} \tag{1}$$

where $\eta \in \Upsilon \subseteq \mathbb{R}^q$ is the model parameter and $Z{:}\mathcal{X} \to \mathbb{R}^q$ are statistics based on the adjacency matrix [3, 4]. There is an extensive literature on descriptive statistics for networks [5, 6]. 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 $Z(X)$ [7]. 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 $c(\eta, \mathcal{X})$ is the normalizing function that ensures the distribution sums to one: $c(\eta, \mathcal{X}) = \sum_{y \in \mathcal{X}} \exp\left[\eta \cdot Z(y)\right]$. This factor varies with both η and the support $\mathcal{X}$ 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 [3]. 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 [3] that the class of homogeneous Markov undirected networks is exactly those having the *degree parameterization*:

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

$$N_\Delta(x) = \frac{1}{6} \sum_{i,j,k} x_{ij} x_{jk} x_{kl},$$

where $d_k(x)$ counts the proportion of individuals with degree k and $N_\Delta(x)$ 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 $Z_k(x) = d_k(x)$, $k = 1, \ldots, n-1$, $Z_n = N_\Delta(x)$, $q = n, \eta \in \Upsilon = \mathbb{R}^n$. This parameterization has the advantage that it is directly interpretable in terms of concurrency of partnerships

(i.e. $d_m(x)$ for $m > 0$ is the proportion of individuals with exactly m concurrent partners).

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

$$C(x) = \frac{3N_\Delta(x)}{N_3(x)}$$

where $N_3(x)$ is the number of connected triples of nodes (i.e., 2−stars [3]). This describes the proportion of complete triads in the networks out of a total number of possible triads.

In the remainder of this paper we focus on the following novel model

$$\log\left[P_\theta(X = x)\right] = \eta(\phi) \cdot d(x) + \nu C(x) - \log c(\phi, \nu, \mathcal{X}), \tag{2}$$

where $x \in \mathcal{X}, \theta = (\phi, \nu), \Theta \subset \mathbb{R}^n$, $d(x) = \{d_1(x), \ldots, d_{n-1}(x)\}$. 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 [7].

This model is a curved exponential family if Θ is a smooth curve in $\Upsilon = \mathbb{R}^n$ [8, 9]. 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 [10]. 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 [8]. 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 Erdós[11].

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

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

where x_{ij}^+ is the network with a partnership between i and j and the rest of the network equal to x_{ij}. x_{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:

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

Thus ρ is the mean clustering coefficient over networks in $\mathcal{X}$. 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 $\mathcal{X}$.

The two parameterizations represent the same model class [9]. Translating between equivalent parameters is achieved using the MCMC algorithm given in Section 3 [9, 8].

2.2 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)

$$P_\theta(K = k) = \mathbb{E}_\theta\left[d_k(X)\right] \qquad k = 0, \ldots, n-1$$

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

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

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., [12]). 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 [13, 14]. The limiting distributions of this mechanism can be characterized by long tails.

2.3 Simple Preferential Attachment Models

A mechanism that has been suggested for the formation of power-law social networks is preferential attachment [15, 16, 2]. This and related stochastic processes have a long history in applied statistics [17, 18, 19]. Consider a population of r people in in which (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

previously sexually inactive person, and (2) otherwise the probability that the $r+1$st partnership will be to a person with exactly k partners is proportional to $kf(k|r)$, where $f(k|r)$ is the frequency of nodes with exactly k connections out of the r total links in the population. The limiting distribution of this process is known as the Waring distribution [19]. The Yule distribution discussed by [17] and used by [20] 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 [21] is:

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

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:

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

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. Both the Waring and the Yule distributions have been re-discovered, apparently without awareness of their historical antecedents, by [22] and [23] respectively in the context of modeling growth of the Internet.

3 Generating Random Networks with Specified Structure

Markov Chain Monte Carlo (MCMC) algorithms for generating from the model (1) have a long history and been well studied (see [24] for a review). The basic idea is to generate a Markov chain whose stationary distribution is given by equation (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 $X_{ij} = 1$ or $X_{ij} = 0$ at the next step of the chain. One way to do this is using Gibbs sampling, whereby the new value of X_{ij} is sampled from the conditional distribution of X_{ij} conditional on the rest of the network. Denote "the rest of the network" by X^c_{ij}. Then $X_{ij}|X^c_{ij} = x^c_{ij}$ has a Bernoulli distribution, with odds given by

$$\frac{P(X_{ij} = 1|X^c_{ij} = x^c_{ij})}{P(X_{ij} = 0|X^c_{ij} = x^c_{ij})} = \exp\{\eta \cdot \Delta(Z(x))_{ij}\},$$

where $\Delta(Z(x))_{ij}$ denotes the difference between $Z(x)$ when x_{ij} is set to 1 and $Z(x)$ when x_{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 x_{ij}. This proposal is accepted with probability $\min\{1, \pi\}$, where

$$\begin{aligned}\pi =& \frac{P(X_{ij} = 1 - x_{ij} | X^c_{ij} = x^c_{ij})}{P(X_{ij} = x_{ij} | X^c_{ij} = x^c_{ij})} \\ =& \begin{cases} \exp\{\eta \cdot \Delta(Z(x))_{ij}\} & \text{if } x_{ij} = 0; \\ \exp\{-\eta \cdot \Delta(Z(x))_{ij}\} & \text{if } x_{ij} = 1. \end{cases}\end{aligned} \tag{6}$$

The vector $\Delta(Z(x))_{ij}$ used by these MCMC schemes is often much easier to calculate directly than as the difference of two separate values of $Z(x)$. For instance, if one of the components of the $Z(x)$ vector is the total number of partnerships in the network, then the corresponding component of $\Delta(Z(x))_{ij}$ 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 x_{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 [25, 26, 27]. More sophisticated MCMC schemes have been developed and are a topic of ongoing research [8].

A variant of this algorithm proceeds in two steps:

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

$$P_\nu(X = x | d_k(X) = d_k) = \frac{\exp[\nu C(x)]}{c(\nu, d_k, \mathcal{X})} \quad x \in \mathcal{X}(d_k)$$

where $\mathcal{X}(d_k) = \{x \in \mathcal{X} : d_k(x) = d_k\}$.

The first generates individual degrees from an arbitrary distribution, and the second generates networks condition on those degrees. Note that the structure of the exponential family in (1) ensure that the samples are from the correct distribution [7]. 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 be a draw from a random distribution) or using sequential importance sampling. An important property of this the second step is the independence of the distribution from ϕ. It is a simple parameter distribution depending only on ν [7].

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

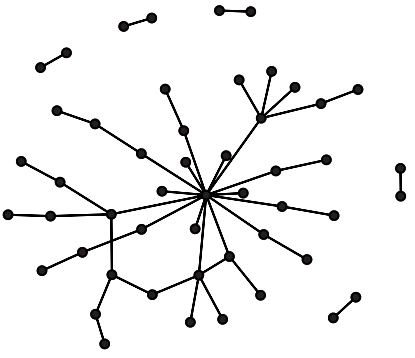

Fig. 1. An example network generated from model (2) with $n = 50$ and degree distribution draw from the Yule model (equation 4) with scaling exponent $\phi_2 = 3$. The random network is drawn from the model with mean clustering coefficient $\rho = 3\%$. The network has clustering coefficient $C(x) =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

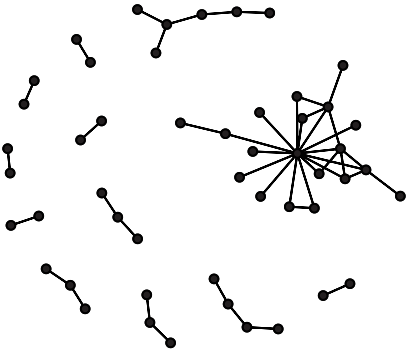

Fig. 2. An example network generated from model (2) with $n = 50$ and degree distribution draw from the Yule model (equation 4) with scaling exponent $\phi_2 = 3$. The random network is drawn from the model with mean clustering coefficient $\rho = 15\%$. The network has clustering coefficient $C(x) =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

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%. Fig. 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.

As an second application we generate a network model for $n = 1000$ nodes with the same degree distribution ($\phi_2 = 3$). A realization of this model is given in Fig. 3. The clustering coefficient for this network is 2%. Fig. 4 is a realization from the model with mean clustering coefficient chosen to be $\rho = 15\%$ (corresponding

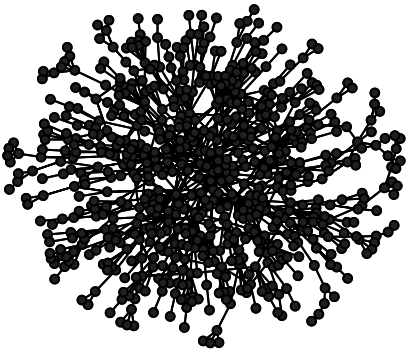

Fig. 3. A random network from the model with $n = 1000$ and the same Yule degree distribution with $\phi_2 = 3$. The largest component is visualized. The network is drawn from the model with mean clustering coefficient $\rho = 3\%$. The realized network has clustering coefficient $C(x) = 1\%$.

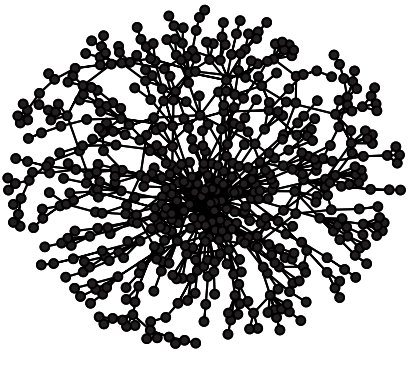

Fig. 4. A random network from the model with $n = 1000$ and the same Yule degree distribution with $\phi_2 = 3$.. The largest component is visualized. The network is drawn from the model with mean clustering coefficient $\rho = 15\%$. The realized network has clustering coefficient $C(x) = 14\%$.

to a clustering parameter of $\nu = 27$.) The elongated nature of the resulting network is apparent as is the centralization of the clustering.

4 Statistical Inference for Network Models

As we have specified the full joint distribution of the network through (1), we choose to conduct inference within the likelihood framework [28, 24]. For economy of notation, in this section, we use ϕ to represent either η in (1) or the curved exponential family form (ϕ, ν) in (2). Differentiating the loglikelihood function:

$$\ell(\phi; x) \equiv \log\left[P_\eta(X = x)\right] = \eta(\phi) \cdot Z(x) - \log\left[c(\phi, \mathcal{X})\right] \tag{7}$$

shows that the maximum likelihood estimate $\hat{\phi}$ satisfies

$$\nabla \ell(\hat{\phi}) = \nabla \eta(\hat{\phi}) \cdot \left[Z(x_{\text{obs}}) - \mathrm{E}_{\,\eta(\hat{\phi})}\, Z(X) \right], \tag{8}$$

where $\nabla \eta(\phi)$ is the $p \times q$ matrix of partial derivatives of η with respect to ϕ and $Z(x_{\text{obs}})$ is the observed network statistic. We may search for a solution to equation (8) using an iterative technique such as Newton-Raphson; however, the exponential family form of the model makes the Fisher information matrix

$$I(\phi) = \nabla \eta(\phi) \cdot \left[\mathrm{Cov}_{\eta(\phi)}\, Z(X) \right] \nabla \eta(\phi) \tag{9}$$

easier to calculate than the Hessian matrix of second derivatives required for Newton-Raphson. For more about equations (8) and (9), see [8] The method of Fisher scoring is an iterative method analogous to Newton-Raphson except that the negative Fisher information is used in place of the Hessian matrix.

Direct calculation of the log-likelihood by enumerating $\mathcal{X}$ is infeasible for all but the smallest networks. As an alternative, we approximate the likelihood equations (8) by replacing the expectations by (weighted) averages over a sample of networks generated from a known distribution. This procedure is described in [24]. To generate the sample we use the MCMC algorithm of Section 3.

5 Application to a Protein-Protein Interaction Network

As an application of these methods, we fit the model to a biological network of protein-protein interactions found in cells. By interact is meant that two amino acid chains were experimentally identified to bind to each other. The network is for *E. Coli* and is drawn from the "Database of Interacting Proteins (DIP)" [29]. The DIP database lists protein pairs that are known to interact with each other. The dataset we use is `Ecoli20050403`. We have chosen *E. Coli* as it is well studied and this will minimize the number of false-negative interactions (that is, two proteins that interact but are not in the database). For simplicity we focus on proteins that interact with themselves and have at least one other interaction. We do not represent the self-interactions as part of the network. This results in a network in Figure 5 with 108 proteins and 94 interactions.

We consider the model (2) with a clustering coefficient term and the degree distribution model by a preferential attachment process (the Yule distribution with scaling exponent ϕ). We choose the Yule as it represents the simple version of preferential attachment that is common in the literature. The estimates are given in Table 1. They are derived using the algorithm in Section 4.

The estimate of the preferential attachment scaling decay rate of about three suggests that the network is close to the so-called "scale-free" range (that is, $\phi \leq 3$). We note that the standard errors are based on the curvature of the estimated log-likelihood and approximations to the sampling distribution based on asymptotic arguments require non-standard justifications. In this case the standard approximation to the sampling distribution can be shown to be valid using a parametric bootstrap. The standard error of the scaling rate indicates

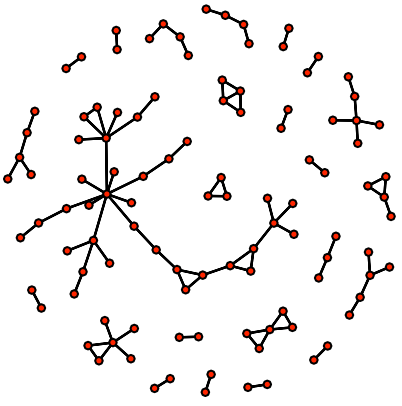

Fig. 5. A protein - protein interaction network for *E. Coli*. The nodes represent proteins and the partnerships indicate that the two proteins are known to interact with each other.

Table 1. MCMC maximum likelihood parameter estimates for the protein-protein interaction network

Parameter	est.	s.e.
Scaling decay rate (ϕ)	3.034	0.3108
Correlation Coefficient (ν)	1.176	0.1457

some uncertainty in the determination of the rate. However the parameter of the correlation coefficient is very positive. This indicates strong clustering (given the degree sequence) and hence so-called "small world" behavior in the network. Thus, this model provides a statistical valid means to test for small-world characteristics of a network using the statistics commonly used to characterize small-world networks.

Finally, we can test if the network is generated by this preferential attachment model. If preferential attachment among proteins generated this network then the parameter ν of the clustering coefficient will be zero. However we see that the estimate is positive. We can test this more rigorously by comparing the log-likelihood values for the maximum likelihood fit in Table 1 to the model where ν is constrained to be zero. The change in the log-likelihood is 52.3, so that the change in deviance is 104.6. This indicates that deviation from the preferential attachment model is statistically significant, as can be verified by a parametric bootstrap of the change in deviance.

6 Discussion

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.

The model form (1) is very general, and can incorporate general social structure [3, 30, 9, 8]. 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 [31]. This model is the model (2) with $\rho = 0$ and $\mathcal{X}$ 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 [32, 33].

References

[1] Newman, M.E.J.: Spread of epidemic disease on networks. Physical Review E **66**(1) (2002) art. no.–016128

[2] Dezsó, Z., Barabási, A.L.: Halting viruses in scale-free networks. Physical Review E **65** (2002) art. no. 055103

[3] Frank, O., Strauss, D.: Markov graphs. Journal of the American Statistical Association **81**(395) (1986) 832–842

[4] Handcock, M.S.: Degeneracy and inference for social network models. In: Paper presented at the Sunbelt XXII International Social Network Conference in New Orleans, LA. (2002)

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

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

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

[8] Hunter, D.R., Handcock, M.S.: Inference in curved exponential family models for networks. Journal of Computational and Graphical Statistics (2006) to appear

[9] Handcock, M.S.: Assessing degeneracy in statistical models of social networks. Working paper #39, Center for Statistics and the Social Sciences, University of Washington (2003)

[10] Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. Physical Review E **64** (2001) 026118

[11] Bollobas, B.: Random Graphs. Academic Press, NY, USA (1985)

[12] Newman, M.E.J.: The structure and function of complex networks. SIAM Review **45**(2) (2003) 167–256

[13] Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439) (1999) 509–512

[14] Pastor-Satorras, R., Vespignani, A.: Epidemic dynamics and endemic states in complex networks. Physical Review E **63**(6) (2001) art. no.–066117

[15] Albert, R., Barabási, A.L.: Topology of evolving networks: Local events and universality. Physical Review Letters **85**(24) (2000) 5234–5237

[16] Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Åberg, Y.: The web of human sexual contacts. Nature **411**(6840) (2001) 907–908
[17] Simon, H.: On a class of skew distribution functions. Biometrika **42**(3/4) (1955) 435–440
[18] Kendall, M.: Natural law in the social sciences: Presidential address. Journal of the Royal Statistical Society Series A-Statistics in Society **124**(1) (1961) 1–16
[19] Irwin, J.: The place of mathematics in medical and biological statistics. Journal of the Royal Statistical Society Series A (General) **126**(1) (1963) 1–45
[20] Jones, J., Handcock, M.S.: An assessment of preferential attachment as a mechanism for human sexual network formation. Proceedings of the Royal Society of London, B **270** (2003) 1123–1128
[21] Johnson, N., Kotz, S., Kemp, A.: Univariate discrete distributions. 2nd edn. Wiley series in probability and mathematical statistics. Wiley, New York (1992)
[22] Levene, M., Fenner, T., Loizou, G., Wheeldon, R.: A stochastic model for the evolution of the web. Computer Networks **39**(3) (2002) 277–287
[23] Dorogovtsev, S.N., Mendes, J.F.F., Samukhin, A.N.: Structure of growing networks with preferential linking. Physical Review Letters **85**(21) (2000) 4633–4636
[24] Geyer, C.J., Thompson, E.A.: Constrained monte carlo maximum likelihood calculations (with discussion). Journal of the Royal Statistical Society, Series B **54** (1992) 657–699
[25] Snijders, T.A.B.: Markov chain monte carlo estimation of exponential random graph models. Journal of Social Structure **3**(2) (2002)
[26] Handcock, M.S.: Progress in statistical modeling of drug user and sexual networks. Manuscript, Center for Statistics and the Social Sciences, University of Washington (2000)
[27] Snijders, T.A.B., Pattison, P., Robins, G.L., Handcock, M.S.: New specifications for exponential random graph models. Sociological Methodology **36** (2006) to appear
[28] Besag, J.: Statistical analysis of non-lattice data. The Statistician **24** (1975) 179–95
[29] Xenarios, I., Salwinski, L., Duan, X.J., Higney P., Kim, S., Eisenberg, D.: Dip, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res **28** (2002) 289–291
[30] Strauss, D., Ikeda, M.: Pseudolikelihood estimation for social networks. Journal of the American Statistical Association **85** (1990) 204–212
[31] Newman, M.E.J.: Assortative mixing in networks. Physical Review Letters **89**(20) (2002) art no. – 208701
[32] Morris, M.: 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, DC (2003) 174–186
[33] Handcock, M.S.: 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, DC (2003) 229–240