Comparison of Maximum Pseudo Likelihood and Maximum Likelihood Estimation of Exponential Family Random Graph Models [1]

Working Paper no. 74
Center for Statistics and the Social Sciences
University of Washington

Marijtje A.J. van Duijn
University of Groningen, Groningen, The Netherlands

Krista Gile
University of Washington, Seattle

Mark S. Handcock
University of Washington, Seattle

April 29, 2007

[1] Marijtje A.J. van Duijn is Assistant Professor of Statistics and Measurement Theory, Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands. E-mail: m.a.j.van.duijn@rug.nl; Web:http://ppswmm.ppsw.rug.nl/~vduijn. Her work on this project was supported by a Fulbright visiting scholarship 2005-2006 to the Center of Statistics and the Social Sciences, University of Washington. Krista Gile is a graduate student, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4332. E-mail: kgile@stat.washington.edu; Web: http://www.stat.washington.edu/kgile/; Mark S. Handcock is professor of Statistics, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4322. E-mail: handcock@stat.washington.edu; Web: http://www.stat.washington.edu/handcock. Corresponding author. Tel.: +1-206-221-6930; Fax: +1-206-221-6873

Abstract

The statistical modeling of social network data is difficult due to the complex dependence structure of the tie variables. Statistical exponential families of distributions provide a flexible way to model such dependence. They enable the statistical characteristics of the network to be encapsulated within an exponential family random graph (ERG) model. For a long time, however, likelihood-based estimation was only feasible for ERG models assuming dyad independence. For more realistic and complex models inference has been based on the pseudo-likelihood. Recent advances in computational methods have made likelihood-based inference practical, and comparison of the different estimators possible.

In this paper, we compare the bias, standard errors, coverage rates and efficiency of maximum likelihood and maximum pseudo-likelihood estimators. We also propose an improved pseudo-likelihood estimation method aimed at reducing bias. The comparison is performed using simulated social network data based on two versions of an empirically realistic network model, the first representing Lazega's law firm data and the second a modified version with increased dependency. We consider estimation of both the natural parameters and the mean-value parameters.

The results clearly show the superiority of the likelihood-based estimators over those based on pseudo-likelihood. The use of the mean value parameterization provides insight into the differences between the estimators and when these differences will matter in practice.

KEY WORDS: exponential family random graph model; networks; maximum likelihood estimation; maximum pseudo likelihood estimation; dyad dependence; mean value parameterization; Markov Chain Monte Carlo

1 Introduction

Maximum likelihood estimation of exponential family random graph (ERG) models is complicated because the likelihood function is difficult to compute for models and networks of reasonable size (e.g., networks with 30 or more actors and models of dyad dependence). Until recently inference for ERG models has been almost exclusively based on a local alternative to the likelihood function referred to as the *pseudo-likelihood* (Strauss and Ikeda, 1990). This was originally motivated by (and developed for) spatial models by Besag (1975), and extended as an alternative to maximum likelihood estimation for networks (Frank and Strauss, 1986; Strauss and Ikeda, 1990; Frank, 1991), (see also Wasserman and Pattison, 1996; Wasserman and Robins, 2004; Besag, 2000). The computational tractability of the pseudo-likelihood function makes it an attractive alternative to the full likelihood function.

In recent years much progress has been made in likelihood-based inference for ERG models by the application of Markov Chain Monte Carlo (MCMC) algorithms (Geyer and Thompson, 1992; Crouch et al., 1998; Corander et al., 1998, 2002; Handcock, 2002; Snijders, 2002; Hunter and Handcock, 2006). At the same time we have gained a far better understanding of the problem of degeneracy (Snijders, 2002; Handcock, 2003; Snijders et al., 2006; Robins et al., 2007).

Since ERG models are within the exponential family class, the properties of their maximum likelihood estimator (MLE) have been studied, although little is available on their application to network models. Little is known about the behavior of the maximum pseudo-likelihood estimator (MPLE), and how these compare with the MLE. Corander et al. (1998) investigate maximum likelihood estimation of a specific type of exponential family random graph model (with the number of two-stars and triangles as sufficient statistics) and compare them to MPLE in two ways. First, they consider small graphs with fixed sufficient statistics and compare the actual MLE (determined by full enumeration), estimated MLE, and MPLE estimates, and conclude that the MPLE estimates appear biased. They also consider graphs with fixed edge counts generated by a model with known clustering parameter. They estimate this known parameter using MLE and MPLE and conclude that outside the unstable region of the MLE, MPLE is more biased than MLE where this difference is smaller for larger networks (40-100 nodes). Wasserman and Robins (2004) argue that the MPLE is intrinsically highly dependent on the observed network and, consequently, may result in substantial bias in the parameter estimates for certain networks.

Lubbers and Snijders (2006) investigated the behavior of MPLE and MLE in several specifications of the ERG model. Their work is not a simulation study but can be characterized as a meta analysis of a large number of same gender social networks of adolescents.

Lubbers and Snijders (2006) conclude that although the behavior of MPLE and MLE can be quite divergent in a particular social network, the overall conclusion obtained in the meta analysis of all models is not seriously affected by the estimation method.

Approaches to avoid the problem of ERG model degeneracy and associated estimation problems include the use of new network statistics (Snijders et al., 2006; Hunter and Handcock, 2006) in the ERG model specification. Handcock (2003) proposed mean value parameterization as an alternative to the common natural parameterization to enhance the understanding of the degeneracy problem.

Because the pseudo-likelihood can be expected to ignore at least part of the dependence structure of the social network, it is generally assumed that inference based on the pseudo-likelihood is problematic. Of particular concern is the underestimation of the standard errors (cf. Wasserman and Robins, 2004). Moreover, it has the undesirable property of sometimes resulting in infinite estimates (manifested by reported estimates that have numerically large magnitude). Handcock (2003) shows that, under certain conditions, if the MPLE is finite, it is also unique. Corander et al. (1998) however, found considerable variability and bias in the MPLE of the effects of the number of 2-stars and the number of triangles in relatively small undirected networks with a fixed number of edges. They also showed that the bias and mean-squared error of MLE are associated with the size of the parameter values, as is typical for parameter configurations where model degeneracy may become a problem.

For dyad independence models the likelihood and pseudo-likelihood functions coincide (if the pseudo-likelihood function is defined at the dyad-level). The common assumption then is that the estimates will diverge as the dependence among the dyads increases. It is also commonly presumed that differences between the estimators will be smaller for actor (or dyadic) covariate effects than for structural network effects, such as the number of transitive triads. Moreover, if the dependence in the network is relatively low, the MPLE may be a reasonable estimate. In analyzes of social network data with many possible (actor) covariates, the easily and quickly obtained MPLE may provide a good starting point.

The purpose of this paper is to study and compare the properties of MPLE and MLE in a case study of a known well-behaved model analyzing part of the data collected by Lazega (2001) and a derived model with higher dyad dependence. We also propose and analyze the properties of a modified MPLE designed to reduce bias in generalized linear models (Firth, 1993). A second goal of the paper is to investigate the difference between the natural and mean-value parameterization of the ERG model. After introducing the exponential family random graph model in both parameterizations, the study design is presented in detail in the third section. The results of the comparison are given in Section 4. The paper concludes with a discussion and recommendations.

2 Exponential Family Random Graph 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 $\mathcal{Y}$ denotes the set of all possible networks on the given n individuals. The multivariate distribution of Y can be parameterized in the form:

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

where $\eta \in \Upsilon \subseteq \mathbb{R}^q$ is the model parameter and $Z{:}\mathcal{Y} \to \mathbb{R}^q$ are statistics based on the adjacency matrix (Frank and Strauss, 1986; Handcock, 2002).

This model is an exponential family of distribution with natural parameter η and sufficient statistics $Z(Y)$. 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 $c(\eta, \mathcal{Y})$ is the normalizing function that ensures the distribution sums to one: $c(\eta, \mathcal{Y}) = \sum_{y \in \mathcal{Y}} \exp\left[\eta \cdot Z(y)\right]$. This factor varies with both η and the support $\mathcal{Y}$ and is the primary barrier to simulation and inference under this modeling scheme.

ERG models have usually been expressed in their natural parameterization η. Here we also consider the alternative mean value parameterization for the model: $\mu : \Upsilon \to C$ defined by

$$\mu(\eta) = \mathbb{E}_\eta\left[Z(Y)\right] \tag{2}$$

where C is the relative interior of the convex hull of the sample space of $Z(Y)$. In the mean value parameterization, the natural parameter η is replaced by $\mu(\eta)$, as the expected value of the sufficient statistic $Z(Y)$ under the model with natural parameter η. The range of μ, $\mu(\eta)$, is C (Barndorff-Nielsen, 1978). One advantage of the mean value parameterization is that from the researcher's perspective it is actually more "natural" than the η parameterization because it is defined on the scale of network statistics. See Handcock (2003) for details.

2.1 Illustration for the Rényi-Erdős Model

A directed Rényi-Erdős network is generated by an ERG model for n actors with one model term capturing the density D of arcs,

$$P_\theta(Y = y) = \frac{\exp(\theta D(y))}{c(\theta)}$$

with $D(y) = \frac{1}{N}\sum_{i \neq j} y_{ij}$. and $N = n(n-1)$, the number of possible ties in the social network. It is also referred to as the homogeneous Bernoulli model.

In this case, the normalizing constant is

$$c(\theta) = \sum_{s=0}^{N} \binom{N}{s} \exp(\theta s/N) = (1 + \exp(\theta/N))^N$$

The mean value parameterization for the model is:

$$\mu \equiv \mu(\theta) = E_\theta(D(y)) = \frac{\exp(\theta)}{1 + \exp(\theta)},$$

representing the probability that an tie exists from a given actor to another given actor. It follows that

$$\theta = \log(\frac{\mu}{1-\mu}),$$

is the (common) log-odds that a given directed pair have a tie. So, for the Rényi-Erdős model, we find that the natural parameter θ is a simple function of the mean value parameter, and vice versa.

The gradient or rate of change in θ as a function of μ is $[\mu(1-\mu)]^{-1}$, which is unbounded as the probability approaches 0 or 1. The rate of change in μ as function of θ is equal to $\exp(\theta)/(1 + \exp(\theta))^2$ which can be re-expressed as $\mu(1 - \mu)$, the variance of the number of arcs under the binomial distribution with constant tie probability μ. Thus, the rate of change in the mean value parameterization is bounded between zero and one-quarter and is a (quadratic) function of the network density.

Although it is only in special cases that such a clear relation between the natural and mean value parameterization exists, it can be helpful in understanding the differences between them.

For any given model, both parameterizations can be considered simultaneously. The issues raised by each are similar to those raised by the choice of parameterization for log-linear analysis: log-linear verses marginal parameterizations. See Agresti (2002), section 11.2.5 for details.

2.2 Inference for Exponential Family Random Graph Models

As we have specified the full joint distribution of the network through (1), it is natural to conduct inference within the likelihood framework (Besag, 1975; Geyer and Thompson, 1992). For economy of notation, differentiating the loglikelihood function:

$$\ell(\eta; y) \equiv \log\left[P_{\eta,\mathcal{Y}}(Y = y)\right] = \eta \cdot Z(y) - \log\left[c(\eta, \mathcal{Y})\right] \tag{3}$$

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

$$Z(y_{\text{obs}}) = \mathbb{E}_{\hat{\eta}} Z(Y), \tag{4}$$

where $Z(y_{\text{obs}})$ is the observed network statistic.

As can be seen from (3), direct calculation of the log-likelihood by enumerating $\mathcal{Y}$ is infeasible for all but the smallest networks. As an alternative, we can approximate the likelihood equations (4) by replacing the expectations by (weighted) averages over a sample of networks generated from a known distribution. This procedure is described in Geyer and Thompson (1992). To generate the sample we use a MCMC algorithm (Geyer and Thompson, 1992; Snijders, 2002; Handcock, 2002).

Until recently inference for the model (1) has been almost exclusively based on a local alternative to the likelihood function referred to as the *pseudo-likelihood* (Strauss and Ikeda, 1990). This was originally motivated by (and developed for) spatial models by Besag (1975).

Consider the conditional formulation of the model (1):

$$\text{logit}[P_\eta(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)] = \eta \cdot \delta(y_{ij}^c) \qquad y \in \mathcal{Y} \tag{5}$$

where $\delta(y_{ij}^c) = Z(y_{ij}^+) - Z(y_{ij}^-)$, the change in $Z(y)$ when y_{ij} changes from 0 to 1 while the remainder of the network remains y_{ij}^c (See Strauss and Ikeda, 1990).

The pseudo-likelihood for the model (1) is:

$$\ell_P(\eta; y) \equiv \eta \cdot \sum_{ij} \delta(y_{ij}^c) y_{ij} - \sum_{ij} \log\left[1 + \exp(\eta \cdot \delta(y_{ij}^c))\right] \tag{6}$$

Thus the pseudo-likelihood 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 $\delta(y_{ij}^c)$ The value of the MPLE 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 pseudo-likelihood 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 MPLE. While in common use (Wasserman and Pattison, 1996;

Anderson et al., 1999), the statistical properties of pseudo-likelihood estimators for social networks are poorly understood.

As a second alternative we propose a method to reduce the bias of the MPLE. The method was originally proposed by Firth (1993) as a general approach to reducing the asymptotic bias of maximum likelihood estimates by penalizing the likelihood function. The bias-corrected pseudo-likelihood for the model (1) is then defined as:

$$\ell_{BP}(\eta; y) \equiv \ell_P(\eta; y) + \frac{1}{2}\log|I(\eta)| \tag{7}$$

where $I(\eta)$ denotes the expected Fisher information matrix for the formal logistic model underlying the pseudo-likelihood evaluated at η. We refer to the estimator that maximizes $\ell_{BP}(\eta; y_{obs})$ as the maximum bias-corrected pseudo-likelihood estimator (MBLE). Heinze and Schemper (2002) showed that Firth's method is particularly useful in rare-events logistic regression where infinite parameter estimates may result because of so called (quasi-) separation, the situation where successes and failures are perfectly separated by one covariate or by a linear combination of covariates. This is the manifestation of the computational degeneracy discussed by Handcock (2003).

Computationally, inference under the mean value parameterization is similar to inference under the natural parameterization. While the point estimator is trivial, obtaining high quality measures of uncertainty of the estimator appears to require a MCMC procedure. Note, however, that they can be directly obtained as a byproduct in the MCMC estimation procedure used to estimate the natural parameters.

3 Study design

Given the limited knowledge of the relative behavior of the MLE and the two maximum pseudolikelihood procedures, we compare them in multiple ways. We investigate bias and efficiency in terms of mean-squared error of the natural and mean value parameter estimates. We also compare the estimates of the standard errors, and the coverage properties of the nominal Wald confidence intervals based on the estimates.

We aim to consider many characteristics of the procedures in depth for a specific model, rather than directly compare the point estimates over many data sets. While the latter approach has value, as demonstrated by Lubbers and Snijders (2006), fixing on a model used to represent a realistic data-set enables the study to focus on the many characteristics of the procedures themselves (rather than just their point estimates). This limits the generalizability of this study in terms of range of models, but increases the generalizability in terms of the range of characteristics of the procedures.

The Lazega (2001) undirected collaboration network of 36 law firm partners is used as the basis for the study. The first step is to consider a well fitting model for the data. We focus on one with seven parameters. Typical for the ERG model are the structural parameters, related to network statistics, here the number of edges (essentially the density) and the geometrically weighted edgewise shared partner statistic (denoted by GWESP), a measure of the transitivity structure in the network. Two nodal attributes are used: seniority (ranknumber/36) and practice (corporate or litigation). Three dyadic homophily attributes are used: practice, gender (3 of the 36 lawyers are female) and office (3 different locations of different size). This is Model 2 in Hunter and Handcock (2006). The model has been slightly reparameterized by replacing the alternating k-triangle term with the GWESP statistic. The scale parameter for the GWESP term fixed at its optimal value (0.7781). (See Hunter and Handcock, 2006, for details). A summary of the MLE parameters used is given in column two of Table 1. Note that we are taking these parameters as "truth" and considering networks produced by this model.

In the next step 1000 networks are simulated from this choice of the parameters. For these networks, the MLE, MPLE and MBLE are obtained using `statnet` (Handcock et al., 2003), both for the natural parameterization and for the mean value parameterization (see Handcock, 2003). One sampled network has observed statistics residing on the edge of the convex hull of the sample space. This network is computational degenerate in the sense of Handcock (2003). For such networks the natural parameter MLEs and MPLEs are known to be infinite and their values were not included in the numerical summaries.

The mean value parameters are a function of the natural parameters. Their values are estimated by simulating networks from the natural parameter estimates and computing the mean sufficient statistics over those samples. The bias of each procedure is the difference between the mean parameter estimate over 999 samples and the true parameter values from which the networks were sampled. Similarly, the standard deviation of each procedure is simply the standard deviation of the parameter values over all samples. The mean squared error, used to compute relative efficiency, is the mean of the squared difference between the parameter estimates and the true parameters.

Estimates of the standard errors are based on the curvature of the log-likelihood (or log-pseudo-likelihood), which we call the "perceived" standard errors. This is because these are the values formally derived as the standard approximations to the true standard errors from asymptotic arguments that have not been justified for these models. These expressions are typically used in the standard software (Handcock et al., 2003; Boer et al., 2003).[1]

[1]Note that since the sample is generated from the MLE fit to the original network, the "perceived" standard errors from that model should be the same as the "actual" standard errors across our sample. Examining these values (see Table 3) shows that they are, in fact, the same up to MCMC uncertainty.

Furthermore, "perceived" confidence intervals are computed based on the Wald statistic with presumed t distribution on 60 degrees of freedom. These perceived confidence intervals have nominal coverage of 95%.

Based on the geometry of the likelihood of the exponential family models, the perceived covariance of the mean value parameter estimates is computed from the inverse of the perceived covariance of the natural parameters estimates (Barndorff-Nielsen, 1978).

To investigate the change in relative performance with higher dependence, another set of networks with higher dependence is considered. Dependence is conceptualized in terms of the observed GWESP statistic, as compared to its expected value in the dyad independent graph with a GWESP natural parameter of 0. The original network has GWESP statistic 190.3. Fitting the model to the network omitting the GWESP term gives a GWESP mean value parameter of 136.4, implying that 53.9 units of GWESP are induced by the dependence structure of the original network. From this perspective, increasing dependence by $100\times\alpha$ percent can be represented by a model with mean value GWESP parameter $190.3 + \alpha 53.9$ with the mean value parameters of the other terms unchanged.

Increasing the dependence in the network also increases the problem of degeneracy. Doubling the dependence ($\alpha = 1$), and even adding half of the dependence ($\alpha = .5$) both result in degenerate models with an unacceptable proportion of probability mass on very high density and very low density graphs. Therefore, the higher dependency model considered adds one quarter of the dependence in the original ($\alpha = .25$). An additional 1000 networks are sampled from this model, and fit with each of the three methods considered. In this case, two sampled networks were computationally degenerate and were not included in the numerical summaries due to infinite MLEs and MPLEs.

4 Results

The properties of the original model's natural parameter estimates under maximum likelihood (MLE), maximum pseudo-likelihood (MPLE) and maximum pseudo-likelihood with Firth's bias-correction penalty (MBLE) are summarized in Table 1. The bias and standard deviation are presented in percentages of the true natural parameter values. For almost all terms, the MLE bias is largest, followed by MPLE, and then MBLE. The MLE standard deviations, however, are the smallest of the three for all terms, with MBLE slightly smaller than MPLE. The standard deviations are typically larger than the bias, and therefore the efficiency, defined as the ratio of the mean squared error (MSE) of the MPLE and MBLE to the MLE is never larger than 1. The efficiencies of the MPLE and MBLE are lowest for the GWESP parameter, which is meant to capture the dependence in the network. Due to the smaller bias of the MBLE, it outperforms the MPLE in efficiency. In one case, the case of homophily on office, the bias of the MLE is large enough that the relative efficiency of MBLE compared to MLE is 1.

Table 1: Bias and standard deviation of MLE, MPLE, and MBLE of ERG model natural parameters as percentages of true parameter values and efficiency of MPLE and MBLE with respect to MLE

natural	true	bias (percentage)			std. dev. (percentage)			efficiency	
parameter	value	MLE	MPLE	MBLE	MLE	MPLE	MBLE	MPLE	MBLE
structural									
edges	−6.51	−2.9	−3.2	−0.7	10.2	11.4	10.9	0.80	0.94
GWESP	0.90	−5.2	1.8	−0.7	17.4	22.7	22.0	0.64	0.68
nodal									
seniority	0.85	11.3	4.2	1.9	32.2	36.2	35.4	0.87	0.92
practice	0.41	16.5	5.1	2.5	35.5	40.8	39.7	0.91	0.96
homophily									
practice	0.76	−0.7	0.1	−1.5	28.8	30.1	29.3	0.91	0.96
gender	0.70	12.1	9.0	4.0	43.4	49.2	47.0	0.81	0.91
office	1.15	6.3	3.6	1.2	19.7	21.2	20.6	0.92	1.00

Table 2 is the mean value parameterization analog of Table 1. Here, a different bias ranking of the three estimators is consistent for all parameters: The MLE is approximately unbiased, and the MBLE is much better than MPLE. By definition, the MLE of the mean value parameterization is unbiased. The deviations from 0 observed in Table 2 are an indication of the computational uncertainty due to the MCMC algorithm. The much larger bias of the MPLE provides an indication that maximum pseudo likelihood estimation does not perform very well in replicating the original network statistics. The negative bias implies underestimation of the observed network statistics, which is, to a lesser extent, also true for MBLE estimation. The standard deviation of the MLE estimates are also smaller than the other two by a factor of about 2. Consequently, the efficiencies of MPLE and MBLE are quite low with respect to estimating the mean value parameters.

Table 3 provides the mean of the perceived standard errors over the network samples. For the MLE they are slightly smaller than the sampling standard deviations (reported in Table 1), whereas the standard errors for MPLE and MBLE are larger, except for GWESP.

The perceived confidence intervals appear to work rather well for the MLE, with coverage rates quite close to the nominal 95%. The inaccuracy of the standard errors of MPLE and MBLE is apparent in their approximate coverage rates, which are too high for all parameters except the GWESP term whose coverage percentage is too small. This means that the standard errors for the structural dependence term are underestimated while they are overestimated for the nodal and dyadic attribute terms.

Table 2: Bias and standard deviation of MLE, MPLE, and MBLE of mean value ERG model parameters as percentages of true parameter values and efficiency of MPLE and MBLE with respect to MLE

mean value	true	bias (percentage)			std. dev. (percentage)			efficiency	
parameter	value	MLE	MPLE	MBLE	MLE	MPLE	MBLE	MPLE	MBLE
structural									
edges	115.00	0.4	−12.2	−3.6	14.4	29.1	26.4	0.21	0.29
GWESP	190.31	0.6	−13.4	−3.3	19.5	34.4	32.1	0.28	0.37
nodal									
seniority	130.19	0.5	−12.5	−4.0	14.8	29.5	26.7	0.22	0.30
practice	129.00	0.4	−13.1	−6.3	13.9	29.0	25.8	0.19	0.27
homophily									
practice	72.00	0.0	−11.5	−3.7	15.0	29.0	26.3	0.23	0.32
gender	99.00	0.6	−11.9	−3.5	15.3	29.7	27.1	0.23	0.31
office	85.00	0.3	−11.9	−4.4	15.0	29.3	26.3	0.23	0.32

Mean perceived standard errors and coverage rates for the mean value parameter estimates are reported in Table 4. Again, the perceived standard errors for the MLE and actual sampling standard deviations in Table 2 are very close, resulting in coverage percentages close to 95%. The perceived standard errors for the MPLE and MBLE, however, are far too small, at around $\frac{1}{3}$ of the sampling standard deviations, resulting in coverage rates just above 50% for nominal 95% intervals.

Tables 5 through 8 report the results for the ERG model with increased dependence. We discuss these results in comparison to the results for the original model. The bias in the natural parameterization, shown in Table 5 is of similar size as in the original model. The standard deviations of the natural parameter estimates also do not show dramatic changes for all estimation methods. They are slightly decreased for the structural parameter estimates, and slightly increased for the attribute parameters, causing the relative efficiencies of the MPLE and MBLE with respect to the MLE to fall substantially for most parameters, to less than .8 for most MPLE estimates, less than .9 for most MBLE estimates, and a mere .50 and .55 respectively for the GWESP estimates. The perceived standard errors and coverage rates for the natural parameters are similar to those for the original model, with slightly larger perceived standard errors and nearly identical coverage rates. For the MLE parameters the perceived standard errors appear to be too high, although this does not lead to too high coverage rates.

The bias observed in the mean value parameterization, in Table 6 is larger by at least a factor of two, resulting in a considerable bias for both the MPLE and MBLE estimates.

The sampling standard deviations are also larger in the increased dependence model, with about a 50% increase over the original model. With higher bias, and higher standard

Table 3: Perceived standard errors of MLE, MPLE, and MBLE of natural ERG model parameters as percentage of the true parameter values and coverage rates of the perceived confidence intervals.

natural	SE (percentage)			coverage percentage		
parameter	MLE	MPLE	MBLE	MLE	MPLE	MBLE
structural						
edges	9.6	13.2	12.9	94.9	97.5	98.0
GWESP	17.2	13.3	13.0	92.7	74.6	74.1
nodal						
seniority	30.6	44.4	43.9	94.4	97.8	98.0
practice	32.5	47.6	47.0	94.0	98.1	98.6
homophily						
practice	27.4	34.5	34.1	94.8	98.1	98.1
gender	40.1	57.8	57.0	95.8	98.7	98.8
office	18.4	25.4	25.1	94.2	98.1	98.4

deviations, it is not surprising that the relative efficiencies of the MPLE and MBLE estimates are smaller in the model with increased dependence. The pseudo-likelihood estimators do not reach even 25% of the efficiency of the MLE for any of the terms. The perceived standard errors for the mean value parameterizations are larger for the MLE compared to Table 4, but are too small compared to the sampling standard deviations in Table 6. Surprisingly, the perceived standard errors are smaller for both the MPLE and MBLE, as compared to the original model, so even more underestimated. Therefore, the coverage rates drop for all three estimators, MLE to about 85%, while MPLE and MBLE fall to about 30% coverage for a nominal 95% interval.

5 Discussion

This case study comparing the quality of maximum likelihood and two maximum pseudo-likelihood estimators confirms that maximum likelihood estimation out-performs maximum pseudo-likelihood estimation on a number of measures, for structural and covariate effects.

In a dyad independent model, such as the one in this study with the GWESP parameter removed, the MLE and MPLE would be identical, while the MBLE would be a slight modification of these to reduce the bias of the natural parameter estimates. In the full dyad dependent model considered here, the MLE is able to appropriately deal with the dependence term, while the MPLE and MBLE can only approximate the dependence pattern. Therefore, it is not surprising that the MLE out-performs the pseudo-likelihood methods to the greatest degree in the estimation of the GWESP natural parameter, in terms of both

Table 4: Perceived standard errors of MLE, MPLE, and MBLE of mean value ERG model parameters as percentage of the true parameter values and coverage rates of the perceived confidence intervals.

mean value	SE (percentage)			coverage percentage		
parameter	MLE	MPLE	MBLE	MLE	MPLE	MBLE
structural						
edges	13.5	6.9	7.0	93.1	44.9	49.4
GWESP	18.4	12.1	12.2	91.4	56.7	62.7
nodal						
seniority	13.9	7.1	7.2	91.6	45.5	49.0
practice	13.1	7.9	8.0	93.2	51.0	57.9
homophily						
practice	14.1	8.3	8.4	92.6	52.0	57.1
gender	14.4	7.4	7.5	92.0	46.5	51.6
office	14.2	7.6	7.7	92.5	50.2	54.4

efficiency and coverage rates. The inferior performance of the MPLE and MBLE natural parameters for the nodal and dyadic attribute terms results from the dependence between the GWESP estimates and the estimates for other model terms. Greater variability in the GWESP results in greater variability in other parameters. This uncertainty also leads to inflated variance estimates for the other parameters, contributing to inflated coverage rates of nominal confidence intervals, whereas the GWESP perceived standard errors are underestimated, resulting in too low coverage rates. Therefore, the notion that inference based on the pseudo-likelihood is problematic are confirmed, where pseudo-likelihood based tests for the structural parameters tend to be liberal, and for the nodal and dyadic attributes conservative. A similar pattern appears to hold for the Lubbers and Snijders (2006) study (see their Figure 2).

As a transformation of the full set of natural parameters, the mean value parameterization is even more conducive to uncertainty in the dependence term decreasing performance on all terms. In addition, the MLE is constructed to be unbiased in the mean value parameters. Together, these two effects contribute to the drastically superior performance of the MLE on the mean value scale. The pseudo-likelihood methods show about 3 times the mean squared error as the MLE, and the perceived coverage rates of nominal 95% confidence intervals hover around 50%. In the model with increased dependence, these effects are even stronger, with relative efficiency below .25 and coverage rates below 40%.

The MBLE is constructed to correct for the bias of the MPLE on the natural parameter scale. It does, in fact, show the smallest bias for the natural parameter estimates. In the case of the practice homophily term, this correction is helpful enough to give MBLE a mean squared error at least as good as that of the MLE.

Table 5: Bias and standard deviation of MLE, MPLE, and MBLE of ERG model natural parameters as percentages of true parameter values and efficiency of MPLE and MBLE with respect to MLE, with increased dependence

natural parameter	true value	bias (percentage)			std. dev. (percentage)			efficiency	
		MLE	MPLE	MBLE	MLE	MPLE	MBLE	MPLE	MBLE
structural									
edges	−6.96	−2.3	−4.0	−1.0	9.7	11.7	11.1	0.66	0.80
GWESP	1.21	−4.3	2.7	−0.2	14.4	21.1	20.2	0.50	0.55
nodal									
seniority	0.78	11.1	5.9	3.3	36.0	42.2	41.1	0.78	0.83
practice	0.35	18.2	5.9	2.8	39.9	51.4	49.8	0.72	0.77
homophily									
practice	0.76	−1.0	2.1	0.1	30.8	31.8	30.8	0.94	1.01
gender	0.66	11.3	8.1	2.8	46.3	53.3	51.1	0.78	0.86
office	1.08	6.9	3.7	0.9	20.4	23.9	23.1	0.79	0.87

Table 6: Bias and standard deviation of MLE, MPLE, and MBLE of mean value ERG model parameters as percentages of true parameter values and efficiency of MPLE and MBLE with respect to MLE, with increased dependence

mean value parameter	true value	bias (percentage)			std. dev. (percentage)			efficiency	
		MLE	MPLE	MBLE	MLE	MPLE	MBLE	MPLE	MBLE
structural									
edges	115.00	−1.1	−26.5	−13.3	20.3	44.7	43.4	0.15	0.20
GWESP	203.79	−1.3	−27.9	−13.4	24.5	49.0	48.2	0.19	0.24
nodal									
seniority	130.19	−1.2	−26.7	−13.6	21.3	45.0	43.6	0.17	0.22
practice	129.00	−0.9	−27.8	−17.1	17.6	43.1	41.0	0.12	0.16
homophily									
practice	72.00	−1.2	−24.8	−12.4	19.5	44.4	42.9	0.15	0.19
gender	99.00	−1.1	−26.4	−13.4	21.6	45.3	44.1	0.17	0.22
office	85.00	−1.0	−26.9	−14.8	20.2	44.5	42.6	0.15	0.20

Although the main point of this paper is to compare the MLE to the pseudo-likelihood methods, it is worth noting that the MBLE consistently out-performs the MPLE in these analyses. The original intent of the method was to reduce the bias of the natural parameter estimates, and it is successful here. However the MBLE also reduces the bias of the mean value parameter estimates relative to the MPLE. Intuitively this is because of the large bias in the mean value parameterizations of the MPLE, and impact of the penalty term jointly on all parameters.

As an illustration, in Figure 1 the MLE, MPLE and MBLE mean values for the edges are plotted (in deviation from the number of edges in the original network), for the networks sampled from the original network (top two panels) and for the sampled networks with increased dependence (the two panels at the bottom). Given that the MLE is unbiased (apart

Table 7: Perceived standard errors of MLE, MPLE, and MBLE of natural ERG model parameters as percentage of the true parameter values and coverage rates of the perceived confidence intervals, with increased dependence.

natural	SE (percentage)			coverage percentage		
parameter	MLE	MPLE	MBLE	MLE	MPLE	MBLE
structural						
edges	14.9	13.9	13.6	96.4	98.2	98.2
GWESP	19.5	12.4	12.1	94.2	78.8	77.6
nodal						
seniority	47.8	53.2	52.4	95.4	98.4	98.7
practice	49.0	61.4	60.5	95.5	98.4	98.8
homophily						
practice	40.9	37.4	36.8	94.6	97.9	98.0
gender	65.7	66.4	65.3	95.3	98.1	98.8
office	26.9	29.3	28.8	95.1	98.2	98.4

from sampling error), the top left panel demonstrates that although the MPLE and MLE are close for many sampled networks, the MPLE underestimates the edge parameter for many other networks, and that it has a larger variance than the MLE with a far less symmetric distribution. The line in the panel indicates the 75% density region, around the $x = y$ line. The remaining points seem to clutter at the bottom of the panel, indicating that MPLE may result in networks with (extremely) low edges, for non-extreme MLE values. From the top right panel, where MPLE is plotted against MBLE, it is clear that the bias correction of MBLE corrects for some, but not all, of this underestimation. Most of the correction occurs for networks with (extremely) low edge MPLE. The degree of underestimation and of correction are greater in the increased dependence model, shown in the bottom two panels. The density lines in the left panel show a substantively larger cluttering of extremely low MPLE edge parameters. Plots for MBLE vs. MLE (not shown) are very similar to those for MPLE vs. MLE.

The study results are slightly conservative in favor of the MPLE and MBLE as the MLE results include the additional computational uncertainty of the MCMC algorithm to estimate the MLE. In the case of the bias of mean value parameter MLEs, which are known to be 0, computational biases are intentionally left in place to represent the performance of the estimators as they might be used in practice. The one deviation from this principle is the exclusion of three "degenerate" sample networks (with observed sufficient statistics at the edges of their possible range), one from the original model and two from the increased dependence model, from the final analysis.

It is worthwhile noting that computational complexity provides another potential differ-

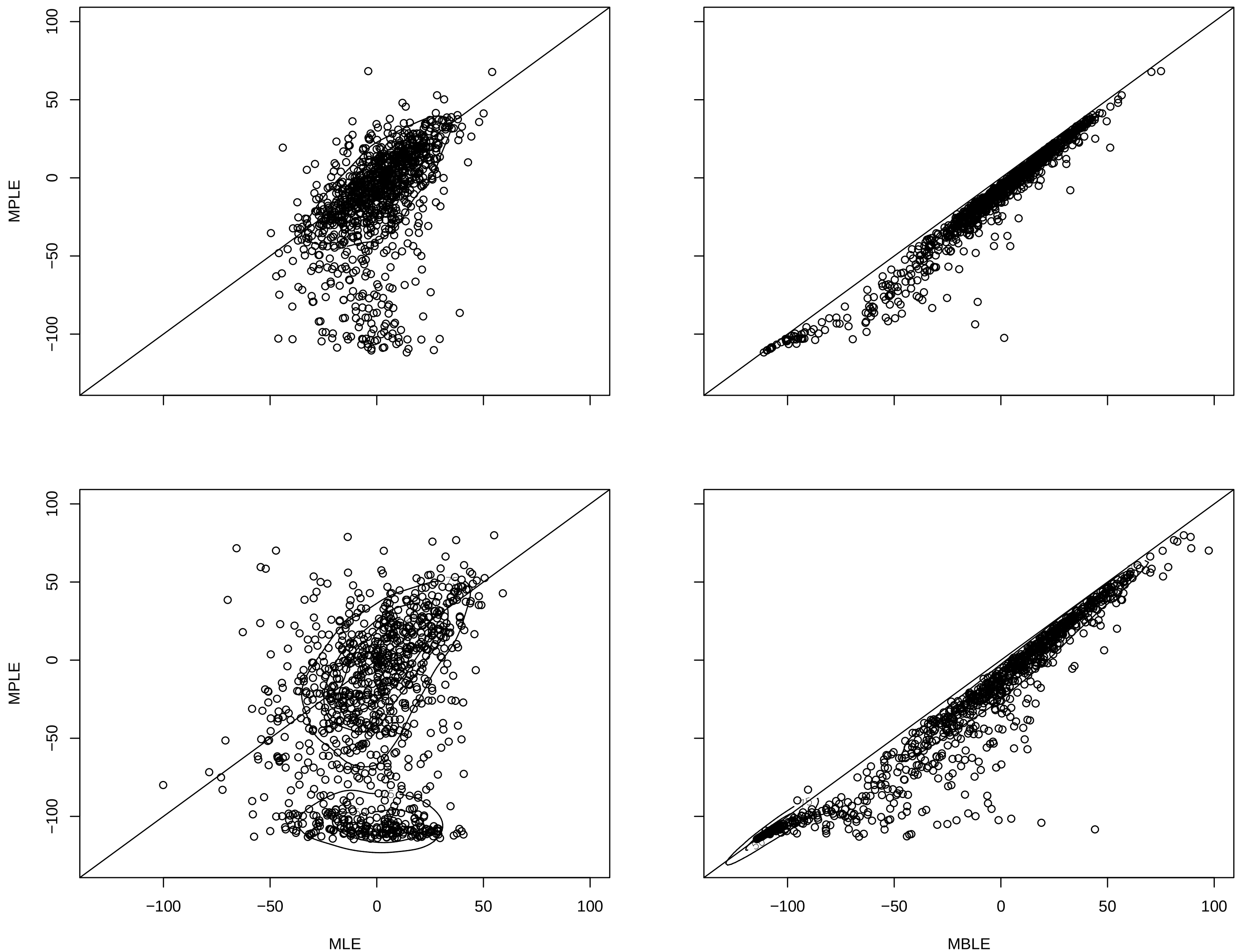

Figure 1: Comparison of error in mean value parameter estimates for edges in original (top) and increased dependence (bottom) models.

Table 8: Perceived standard errors of MLE, MPLE, and MBLE of mean value ERG model parameters as percentage of the true parameter values and coverage rates of the perceived confidence intervals, with increased dependence.

mean value	SE (percentage)			coverage percentage		
parameter	MLE	MPLE	MBLE	MLE	MPLE	MBLE
structural						
edges	17.1	6.5	6.6	85.5	23.8	28.5
GWESP	20.7	10.4	10.5	85.9	31.3	36.6
nodal						
seniority	17.6	6.7	6.8	84.4	22.8	27.6
practice	16.0	7.5	7.6	89.9	35.9	39.3
homophily						
practice	17.2	7.8	7.8	89.7	31.1	37.3
gender	18.0	7.0	7.0	84.8	22.7	28.5
office	17.4	7.2	7.2	87.8	27.0	32.3

ence between partial and maximum likelihood estimation. In fact, this is one reason MPLE estimation has been used for so long. Recent advances in computing power and in algorithms has made MLE a feasible alternative for most applications(Geyer and Thompson, 1992; Snijders, 2002; Hunter and Handcock, 2006). In the Appendix more details about computing time and other computational aspects can be found.

To further investigate differences in the effect of the estimation method on structural and covariate parameters, it would have been good to study a simple triangle model (as Corander et al., 1998, 2002). Unfortunately, this model is degenerate for the Lazega data, providing further evidence that the triangle model may often be too crude to be useful in realistic settings. Corander et al. (1998, 2002) avoid the problem of degeneracy by considering only graphs with a fixed number of edges. The GWESP terms has a similar motivation and fits well on the same data.

We have used standard error estimates from the inverse of the Hessian to compute confidence intervals and coverage rates. We take this approach because these standard error estimates are often used to compute Wald-type confidence intervals and for testing purposes. It is important to remember, however, that we have no asymptotic justification for this approach for models with structural dependence. It might well be the case that this approach leads to worse results in networks with increased dependence, as was found in underestimated perceived standard errors for all mean value parameters (see Table 8). Lack of normality could be another (partial) explanation, in view of the deteriorated coverage rates.

Note that exact testing is an alternative to the Wald approximation (Besag, 2000). To determine an exact $p-$value for a coefficient, for example, simulate from the model conditional on the observed values of the statistics in the model and omitting the target statistic. The $p-$value is then based on the quantile of the observed target statistic among those from the sampled networks. While this approach is feasible, it is usually prohibitively expensive computationally.

Summarizing the main findings of our study, we can make three practical recommendations, in addition to the overall conclusion that it is always better to use MLE than MBLE than MPLE. First, if MLE is not feasible, MBLE is to be preferred over MPLE. Second, if one's main interest is in investigating nodal and dyadic attribute effects, MBLE/MPLE can be useful as a first selection criterion, especially in the natural parameterization, where the bias is reasonably low. We strongly recommend, however, to check any (candidate) final model using MLE, in view of the too liberal tests provided by MPLE/MBLE. For a serious investigation of structural dependence effects, MPLE/MBLE is not recommended. Third, it can be worthwhile to also consider the mean value parameterization to obtain more insight in directly observable and interpretable network characteristics and statistics.

In view of the specificity of the models investigated in this study, we realize that these recommendations may be of limited value. Therefore, as a final recommendation, we encourage further comparison of maximum likelihood estimation to (bias-corrected) pseudo-likelihood estimation in other applications. To this purpose, we have made available the code used in this study on the `statnet` website (see the Appendix for further technical details).

References

Agresti, A. (2002). *Categorical Data Analysis, 2nd ed.* Wiley and Sons, Inc.

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

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

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

Besag, J. (2000). Markov chain monte carlo for statistical inference. Working paper, Center for Statistics and the Social Sciences, University of Washington.

Boer, P., M. Huisman, T. Snijders, and E. Zeggelink (2003). *StOCNET: an open software system for the advanced statistical analysis of social networks* (Version 1.4 ed.). Groningen: ProGAMMA / ICS.

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

Corander, J., K. Dahmström, and P. Dahmström (1998). Maximum likelihood estimation for markov graphs. Research report, Department of Statistics, University of Stockholm.

Corander, J., K. Dahmstrom, and P. Dahmstrom (2002). Maximum likelihood estimation for exponential random graph models. In J. Hagberg (Ed.), *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics; A Festschrift in honour of Ove Frank*, pp. 1–17. Stockholm: University of Stockholm, Department of Statistics.

Crouch, B., S. Wasserman, and F. Trachtenberg (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. (1991). Statistical analysis of change in networks. *Statistica Neerlandica 45*, 283–293.

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

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

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

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

Handcock, M. S., D. R. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris (2003). *statnet: An R package for the Statistical Modeling of Social Networks*. http://csde.washington.edu/statnet.

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

Hunter, D. R. and M. S. Handcock (2006, September). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics 15*(3), 565–583.

Lazega, E. (2001). *The collegial phenomenon: the social mechanisms of cooperation among peers in a corporate law partnership*. Oxford: Oxford University Press.

Lubbers, M. and T. Snijders (2006). A comparison of various approaches to the exponential random graph model a reanalysis of 104 student networks in school classes. Submitted.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Robins, G. L., T. A. B. Snijders, P. Wang, M. S. Handcock, and P. Pattison (2007). Recent developments in exponential random graph (p^*) models. *Social Networks 28*, to appear.

Snijders, T. A. B. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure 3*(2).

Snijders, T. A. B., P. Pattison, G. L. Robins, and M. S. Handcock (2006). New specifications for exponential random graph models. *Sociological Methodology 36*, to appear.

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

Wasserman, S. and P. Pattison (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p^*. *Psychometrika 61*(3), 401–425.

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

Wasserman, S. S. and G. L. Robins (2004). An introduction to random graphs, dependence graphs, and p^*. In J. S. P. Carrington and S. S. Wasserman (Eds.), *Models and Methods in Social Network Analysis*, pp. in press. Cambridge: Cambridge University Press.

A Computational Details

All computations were done using the `statnet` (Handcock et al., 2003) package in the `R` statistical language (R Development Core Team, 2006). The package is freely available for download at `http://www.csde.washington.edu/statnet` and detailed instructions are given there. All estimates are based on the `ergm` function, with various optional arguments. The arguments used are listed in Table 9, and their function is described. Much of this description can be found in Handcock et al. (2003).

Sample networks were fit with two or more successive calls to `ergm`, with arguments listed in the first column of arguments in Table 9. The first two rounds of model-fitting were used for all models. Then, a check for convergence was performed: if the mean value parameter estimate for the `edges` term had error greater than 5 edges, the fit was deemed not sufficiently converged and the third round of model-fitting was repeated until the mean value `edges` term was accurate within 5 edges.

The models are fit using MCMC according to some model parameters, and the `theta0` argument specifies the starting parameter values. Hunter and Handcock (2006) discuss the important role played by starting parameter values in fitting complex ERG models. The `statnet` default is to use the MPLE fit to begin the MLE fit. This approach was used for the samples from the original model. The increased dependence samples, however, were more difficult to fit, so the MPLE estimate proved to be too far from the MLE to lead consistently to converged estimates over large numbers of sampled networks. Two modifications were introduced to address this problem. First, the initial parameter estimates were computed by adding 90% of the MPLE estimate to 10% of the model parameters from the original model. This had the effect of providing some correction in cases where the MPLE estimates were very far from the MLE. Second, the model fit was accomplished using two `ergm` calls. The first was less precise and with smaller sample size, and aimed at producing a rough initial set of parameter estimates closer to the MLE than its original values. The second `ergm` call was started at the estimate produced by the first, and involved a larger sample size in the interest of producing a more precise final set of parameter estimates. Note that these sophistications robustify the estimation of the MLE for the purposes of automation over the 1000 networks, but do not effect the ultimate MLE itself. Based on the results of this study we will use the MBLE estimate as the default in `statnet`.

Given the starting values, the Markov Chain must be created by selecting new proposed sample networks as the basis of the Metropolis-Hastings algorithm. The `proposaltype` argument determines how such new networks are selected. All `ergm` calls in these fits used the `statnet` default `randomtoggle` value of this argument, which selects a dyad at random and evaluates the likelihood ratio of the network with and without a tie in the selected dyad.

The `interval` argument determines the number of Markov Chain steps between successive samples. The `burnin` argument determines the number of initial samples discarded to avoid any possible bias of the original network. And the `MCMCsamplesize` argument determines the total number of samples taken.

Once the sample is completed, the curvature of the MCMC approximation to the likelihood is evaluated and an MCMC-MLE estimate is produced. This estimate was obtained, however, based on a sample from parameter values potentially quite far from the true MLE, so it was potentially inefficient. With a new, improved MLE estimate in hand, it is possible to produce an additional sample based on this estimate to greatly refine the estimate. The `statnet` argument `maxit` does just this: it specifies the number of times an estimate should be produced, with each successive estimate based on a sample from the previous estimate.

The final argument `steplength` modifies the Newton-Raphson optimization of the Monte Carlo approximation to the likelihood to account for the uncertainty in the approximation to the actual likelihood. The value is a between 0 and 1 and indicates how much of the step toward the estimated optimum is taken at each iteration.

Table 9: Arguments to `statnet` function `ergm` used to compute MLE.

Argument	First Round	Second Round	Third Round
theta0	.9 truth + .1 MPLE	prior fit	prior fit
proposaltype	toggle	toggle	toggle
interval	2000	3000	3000
burnin	500000	500000	500000
MCMCsamplesize	5000	10000	5000
maxit	2	2	4
steplength	.7	.7	.5