Conditional Maximum Likelihood Estimation under Various Specifications of Exponential Random Graph Models

Tom A.B. Snijders and Marijtje A.J. van Duijn*

1 Introduction

One among the major contributions by Ove Frank to the statistical analysis of social networks was the introduction, in Frank and Strauss (1986), of the class of *Markov graphs* as a family of distributions for directed and undirected graphs. A random graph is a Markov graph if the number of nodes is fixed (say, at g) and nonincident edges (i.e., edges between disjoint pairs of nodes) are independent conditional on the rest of the graph. Frank and Strauss elaborated on Besag (1974) in their use of the Hammersley-Clifford theorem to characterize Markov graphs as an exponential family of distributions. The model was extended by Frank (1991) and by Wasserman and Pattison (1996) to general exponential families of distributions for graphs, with a focus on directed graphs (digraphs). Wasserman and Pattison (1996) called this family the p^* model. In subsequent work (among others, Pattison and Wasserman, 1999; Robins, Pattison, and Wasserman, 1999) this model was elaborated, mainly using subgraph counts as sufficient statistics.

The exponential family of distributions for a digraph denoted by y, for some vector of sufficient statistics $u = u(y)$, is given by the family of probability functions

*The first author acknowledges stimulating discussions with Philippa Pattison and Garry Robins, made possible by a fellowship at the University of Melbourne (School of Behavioural Sciences); and with Mark Handcock and Martina Morris, made possible by participation in NIH grant 1 R01 DA12831-01A1, "Modeling HIV and STD in Drug User and Social Networks".

$$P_\theta(Y = y) = \exp(\theta' u(y) - \psi(\theta)) \tag{1}$$

where Y is the adjacency matrix of a digraph $y = (y_{ij})_{1 \le i,j \le g}$, with $y_{ij} = 1$ (or 0) indicating the presence (or absence) of an arc from i to j. Note that $y_{ii} = 0$ for all i. The function $\psi(\theta)$ is a norming constant.

Replacing an index by a + sign will be used to indicate summation over this index; e.g., y_{i+} denotes the out-degree of vertex i.

In principle, (1) can represent any probability distribution for digraphs, provided that each digraph has a positive probability. The same formula, restricted to symmetric adjacency matrices y, with $y_{ij} = y_{ji}$ for all i, j, can be used for distributions of undirected graphs. Frank and Strauss (1986) proved that the conditional independence property of the undirected Markov graph is equivalent to model (1) with a sufficient statistic $u(y)$ containing as elements the total number of k-stars for $k = 1, ..., g-1$ and the number of triangles (transitive triads) $\sum_{i<j<k} y_{ij} y_{jk} y_{ik}$.

The general theory of exponential families (e.g. Lehmann, 1983) tells us that $u(y)$ is the canonical sufficient statistic for exponential random graph models and the maximum likelihood (ML) estimator is the solution of the moment equation

$$\mu\left(\widehat{\theta}\right) = u(y) \quad , \tag{2}$$

where $u(y)$ is the observed value and $\mu(\theta)$ is defined as the expected value

$$\mu(\theta) = E_\theta \{u(y)\} \quad .$$

However, this does not help us to find the ML estimate, because for almost all exponential random graph models this expected value is not easily calculated.

Frank and Strauss (1986) discussed methods to estimate the parameters in a Markov graph model where all k-star parameters are 0 for $k > 3$. This leads to a three-parameter exponential family. In the light of the problems associated with explicit calculations, they proposed a simulation-based method to approximate the ML estimate of any one of the three parameters θ_k, given that the other two are fixed at 0. They also proposed a kind of conditional logistic regression method to estimate the full vector θ. This method was elaborated by Strauss and Ikeda (1990), Frank (1991), and Wasserman and

Pattison (1996). It is a pseudolikelihood method which operates by maximizing the so-called pseudologlikelihood defined for digraphs by

$$\ell(\theta) = \sum_{i,j} \ln\left(P_\theta\left\{Y_{ij} = y_{ij} | Y_{hk} = y_{hk} \text{ for all } (h,k) \neq (i,j)\right\}\right) \qquad (3)$$

Although this method is intuitively appealing and easily implemented, the properties of the resulting estimator for exponential graph models are unknown. The pseudolikelihood estimate is not a function of the complete sufficient statistic $u(y)$.

Dahmström and Dahmström (1993) proposed a simulation-based Markov chain Monte Carlo (*MCMC*) method for estimating a single parameter of a Markov graph distribution. This method was extended by Corander, Dahmström and Dahmström (1998) to simulation-based *MCMC* estimation of a multidimensional parameter, following the approach of Geyer and Thompson (1992) to construct Monte Carlo-based algorithms for approximating the Maximum Likelihood estimate.

It was noted by Snijders (2002), however, that simulating the exponential random graph model for a given parameter value, which is a basic step in any *MCMC* estimation method, can have (depending on the parameter values) inherent convergence problems – at least for the commonly used model specifications such as the Markov graph discussed above. The main reason is the bimodal or multimodal shape that the distributions of the sufficient statistics $u_k(Y)$ may have. There are data sets, e.g., where the observed graph density is 0.5, and the ML estimate corresponds to a digraph distribution with a bimodal marginal probability distribution of the density. This bimodal distribution has a probability of almost 0.5 for values very close to 0.0, and a probability of almost 0.5 for values very close to 1.0. This situation indicates a lack of fit between model and data, because for a good fit it is desirable that the main probability mass for the distribution under the ML estimate is concentrated close to the observed outcome. Snijders (2002) proposed a ML estimation algorithm using the Robbins Monro (1951) algorithm to solve the moment equation, but reported that convergence often was insatisfactory.

The statistical problems in estimating exponential random graph models are related to the fact that just one observation Y of the random graph is available; obviously, fitting a potentially bimodally shaped distribution to a single observation is hardly meaningful. A potential way out of these problems is to try to find specifications of the model that do not lead to bimodally shaped distributions. One potential solution is to condition on suitable statistics such as the total number of edges, which will in certain cases take away this bimodally shaped distribution. Another potential solution is based on discussions between the first author, Mark Handcock, Pip Pattison, and Garry Robbins, in which the idea arose that especially the fit of the degree distribution by exponential random graph models may be very poor, and that progress might be obtained by looking for modified model specifications that improve the fit for the distribution of the degrees.

2 Other specifications of exponential random graph models

Conditioning on the total number of edges

If the main problem resides in a bimodal shape for the distribution of the graph density, then an obvious solution is to condition on the total number of edges. Such conditioning was used already by Corander et al. (1998) and was discussed also by Snijders (2002). This paper considers only models with such a conditioning.

Incidental vertex parameters

If one wishes to fit the degree distribution quite precisely, one could include incidental vertex parameters for the in- and out-degrees. For a digraph model this leads to models that are extensions of Holland and Leinhardt's (1981) p_1 model: the vector of sufficient statistics includes all in-degrees and outdegrees, or, equivalently, the exponent $\theta' u(y)$ in the probability (1) includes the term

$$\sum_{i=1}^{n} \left(\alpha_i y_{i+} + \beta_i y_{i+} \right) . \tag{4}$$

This model ingredient was proposed in Wasserman and Pattison's (1996) seminal paper on the p^* distribution, but it was omitted in most more recent publications on this family of distributions, presumably because the Hammersley-Clifford theorem does not give a special motivation for this term and because, traditionally, incidental parameters are disliked in theoretical statistical work.

An exponential random graph model including this term in the exponent comes close to a model containing all k-star counts (up to the highest possible order, $g-1$) except that the latter model will fit the entire in- and outdegree *distribution,* whereas the model including (4) will fit the in- and outdegrees of each of the parameters separately. Admittedly, a model with so many parameters seems not very elegant. But it could be a simple way to fit the degree distribution, and it may have the advantage over a model including k-star counts for high values of k that the parameter estimates are less highly correlated.

The model with vertex parameters can be estimated with or without conditioning on the total number of edges (i.e., the degree sum). The estimation without conditioning turned out to work rather poorly. Therefore we present results only for models with this conditioning.

Conditioning on all degrees

Another option would be not to include the degrees in the sufficient statistic, but to regard the degrees as nuisance parameters and condition on them. This is perfectly in line with the traditional statistical way to deal with nuisance parameters (Lehmann, 1983).

The in- and out-degrees of the random digraph are complete sufficient statistics for the incidental vertex parameters. Therefore, when we condition on the in- and out-degrees, the exponential random graph model still is an exponential family of distributions. If the parameters of the other sufficient statistics are 0, this is the so-called $U \mid \{Y_{i+}\}, \{Y_{+i}\}$ distribution (Holland and Leinhardt, 1975; Wasserman, 1977; Snijders, 1991).

Conditioning on the in- and out-degrees implies that the observed degrees are fitted exactly, but in a trivial way, while it is possible to estimate the other parameters in the model (the parameters for reciprocity, for other subgraph counts, etc.) conditional on these degrees. The perfect fit for the degrees in itself is an advantage; the fact that this trivial way of fitting precludes a parametric estimation of the degree distributions, however, will sometimes be a disadvantage.

The computational advantage is that the outcome space of the exponential random graph is severely restricted, which may (depending on the values of the in- and out-degree) lead to a more stable algorithm.

The appendix gives the principles of simulation-based estimation for random graph models, discussed more extensively by Snijders (2002). The appendix also contains descriptions of the algorithm extensions necessary for the estimation of incidental vertex parameters and for conditioning on the in- and out-degrees.

3 Results for various examples

We investigate whether the above-mentioned model specifications are effective to improve the convergence properties of the simulation-based algorithm for approximating the ML estimate. The Metropolis-Hastings algorithm (Snijders, 2002) is used for simulating the random graphs, conditional on the total number of edges. The estimation methods are applied to a few well-known data sets: Krackhardt's managers data (Krackhardt, 1987) and Freeman's EIES data (Freeman & Freeman, 1980). All data sets considered are directed graphs. The estimations were carried out by SIENA version 1.96 (Snijders and Huisman, 2002). The evaluation whether the outcome θ of the algorithm is a satisfactory solution of the moment equation, is based on estimates for $\mu_k(\widehat{\theta}) = E_{\widehat{\theta}} u_k(Y)$ and $\sigma_k(\widehat{\theta}) = S.D._{\widehat{\theta}} u_k(Y)$. These are calculated in 'Phase three' of the SIENA algorithm. A requirement for good convergence is that the t-values

$$t_k = \frac{u_k(Y) - \mu_k(\widehat{\theta})}{\sigma_k(\widehat{\theta})} \ , \tag{5}$$

where $u_k(Y)$ is the observed statistic, are less than 0.1 in absolute value for all k. The results of the algorithm are stochastic and depend on the initial values. If the algorithm yielded a reasonable but not quite satisfactory result, it was started again from the value found, to try and improve convergence by a good starting value.

In all cases, models are considered that contain the effects of number of ties, number of reciprocated ties, number of transitive triplets, and number of 3-cycles. Exponential random graph model with these four effects and without any conditioning, led to converging estimates only in a few cases. Therefore attention is paid only to conditional models.

3.1 Krackhardt's High-tech Managers

First, the estimation methods were applied to two networks collected by Krackhardt (1987), taken from the data presented in Wasserman & Faust (1994) as Krackhardt's High-tech Managers. The 'advice' and 'friendship' relations in a group of 21 managers are used. For both networks, the model considered contained the effects of number of ties $u_1(y) = y_{++}$, number of reciprocated ties $u_2(y) = \sum_{i,j} y_{ij}y_{ji}$, number of transitive triplets $u_3(y) = \sum_{i,j,k} y_{ij}y_{jk}y_{ki}$, and number of 3-cycles $u_4(y) = \sum_{i,j,k} y_{ij}y_{jk}y_{ki}$ (where summations extend over all non-equal values of the indices). The first of these statistics falls out of the model because of the conditoning.

Table 1: Results for Krackhardt's High-tech Managers friendship relations

		Conditional on number of ties	Conditional on degrees	With vertex parameters
		Parameter estimates (and standard errors)		
Reciprocity		1.56 (0.45)	2.70 (0.59)	2.85
Transitivity		0.35 (0.03)	0.20 (0.10)	−0.07
3-Cycles		−0.56 (0.15)	0.07 (0.24)	0.08
	$u_k(y)$	$\widehat{\mu}_k(\hat{\theta})$ (and $\widehat{\sigma}_k(\hat{\theta})$)		
Reciprocity	23	23.10 (3.10)	23.07 (1.69)	22.86 (2.55)
Transitivity	219	220.50 (43.6)	218.26 (11.3)	218.50 (21.6)
3-Cycles	44	44.32 (12.3)	43.85 (4.75)	43.75 (9.35)

For the friendship network, the density is 0.24, with 102 edges, 23 mutual relations, 219 transitive triples and 44 3-cycles. For the advice network the density is 0.45, with 190 edges, 45 mutual relations, 988 transitive triples and 188 3-cycles. The number of Metropolis-Hastings steps used for generating one network (for one step of the Robbins Monro algorithm) was up to 59,958 for the friendship network, 44,503 for the advice network. (These nonrounded figures follow from how these numbers are determined in SIENA.)

Table 2: Results for Krackhardt's High-tech Managers advice relations

		Conditional on number of ties	Conditional on degrees	With vertex parameters
		Parameter estimates (and standard errors)		
Reciprocity		0.90 (0.38)	2.05 (0.56)	1.83
Transitivity		0.22 (0.02)	0.03 (0.08)	−0.16
3-Cycles		−0.34 (0.05)	0.04 (0.12)	−0.01
	$u_k\,(y)$	$\widehat{\mu}_k(\hat{\theta})$ (and $\widehat{\sigma}_k(\hat{\theta})$)		
Reciprocity	45	45.05 (4.02)	44.94 (1.82)	44.87 (2.83)
Transitivity	988	987.91 (62.9)	988.18 (12.1)	988.19 (30.1)
3-Cycles	188	187.6 (29.5)	188.0 (8.6)	187.2 (17.6)

Convergence is good for all models presented here. A further discussion is given in the final section.

3.2 Freeman's EIES Researchers

Second, the estimation methods were applied to two networks collected by Freeman & Freeman (1980), known as the EIES data, also taken from Wasserman & Faust (1994). Acquaintanceship among a group of researchers is recorded at two time points, one before, and one seven months after the introduction to the then novel phenomenon of communication by computer. Acquaintanceship was recorded on a five-point scale, which we dichotomized, defining the relation as being a friend or close friend. Of the 32 researchers involved, information is available on the number of citations in the year of the data collection, and on the primary discipline. These attributes were dichotomized, defining two equal-sized groups of much and little cited authors (with at least, respectively fewer than 12 citations), and defining a group of sociologists and a mixed group consisting of researchers educated in anthropology, psychology, communication, statistics and mathematics.

For both time points, the model considered contained, next to the network effects also used for Krackhardt's data, the dyadic covariates of dissimilarity with respect to either attribute as defined in SIENA (see Snijders & Huisman, 2002). The tables below report results for the similarity variable defined as minus the dissimilarity variable.

Table 3: Results for Freeman's *EIES* Researchers' friendship at time 1

		Conditional on number of ties	Conditional on degrees	With vertex parameters
		Parameter estimates (and standard errors)		
Reciprocity		2.61 (0.33)	3.45 (0.48)	4.08
Transitivity		0.35 (0.01)	0.33 (0.07)	0.08
3-Cycles		−0.59 (0.07)	−0.07 (0.19)	0.05
Citations sim.		0.23 (0.12)	0.44 (0.15)	0.44
Discipline sim.		0.33 (0.12)	0.37 (0.15)	0.35
	$u_k(y)$		$\widehat{\mu}_k(\hat{\theta})$ (and $\widehat{\sigma}_k(\hat{\theta})$)	
Reciprocity	42	42.11 (4.27)	41.92 (2.21)	42.19 (3.45)
Transitivity	316	322.1 (220)	317.4 (30.7)	312.7 (30.1)
3-Cycles	78	78.6 (49.7)	78.3 (11.3)	76.4 (18.6)
Citations sim.	16.45	15.8 (10.2)	16.8 (7.2)	16.3 (6.5)
Discipline sim.	25.15	24.8 (15.3)	25.3 (6.7)	25.6 (6.1)

For the network at time 1, the density is 0.15, with 152 edges, 42 mutual relations, 316 transitive triples and 78 3-cycles. The observed similarity on citations (after centering around the mean of 0.5) is 16.45; of discipline (mean 0.53) it is 25.15. For the network at time 2 the density is 0.21, with 204 edges, 60 mutual relations, 605 transitive triples and 142 3-cycles. The observed similarity on citations is now 19.29; of discipline (mean 0.53) it is 29.88. The number of Metropolis-Hastings steps used for generating one network was up to 197,306 for the time 1 network, and 156,713 for the time 2 network.

Table 4: Results for Freeman's EIES Researchers' friendship at time 2

		Conditional on number of ties	Conditional on degrees	With vertex parameters
		Parameter estimates (and standard errors)		
Reciprocity		2.67 (0.29)	3.74 (0.45)	4.27
Transitivity		0.29 (0.01)	0.29 (0.06)	0.07
3-Cycles		−0.54 (0.06)	−0.20 (0.15)	−0.06
Citations sim.		0.24 (0.12)	0.30 (0.12)	0.40
Discipline sim.		0.29 (0.10)	0.27 (0.14)	0.27
	$u_k(y)$	$\widehat{\mu}_k(\hat{\theta})$ (and $\widehat{\sigma}_k(\hat{\theta})$)		
Reciprocity	60	59.79 (4.37)	59.93 (2.27)	60.16 (3.63)
Transitivity	605	588.9 (319)	606.7 (31.6)	599.7 (58.8)
3-Cycles	142	135.9 (40.9)	142.9 (11.9)	141.6 (22.8)
Citations sim.	19.29	19.6 (9.0)	18.8 (7.2)	19.4 (6.7)
Discipline sim.	29.88	30.6 (21.0)	29.8 (6.7)	30.0 (6.5)

Note that the parameter estimates for time 2 are strikingly similar to those for time 1. The main difference between time 1 and time 2 is apparently the increase in the number of ties. (One would need another model to investigate this...)

4 Conclusions

The conclusions that can be drawn from the analyses of these four data sets are limited of course, but they do point into a common direction. Note that we succeeded in obtaining ML estimates in unconditional models with triad effects only in exceptional cases.

Conditioning on the total number of edges did lead to satisfactory results. In terms of the t-ratios (5), convergence was good. The relative standard deviations of the two triad counts (transitivity and 3-cycles) in models without vertex parameter are quite high for the EIES data, however. This does imply that the fit of this model is rather mediocre for these data.

The estimation algorithm conditioning on the in- and out-degrees performed even better. However, this conditioning also introduces a substantial difference in the interpretation of the model, as was discussed for conditionally uniform models also in Snijders (1991). The degrees define important restrictions on the possible range of outcomes for many network statistics, and apart from the range they also imply restrictions for the variability as shown by the standard deviations of the subgraph counts in Tables 1−4, which are considerably smaller for the models conditional on the degrees than for those conditional only on the total number of edges. E.g., the 3-cycle effect is significant negative in all examples for the model conditioning only on the total number of edges, but not for the model conditioning on all degrees. In other words, the 3-cycle effect can be 'explained away' by the degree distribution. It depends on subject-matter considerations, however, whether the degrees are considered to be determined by influences extraneous to triadic network effects, which would be required for making such an 'explanation' at all meaningful.

The model with vertex parameters and without conditioning on the degrees (but with conditioning on the total number of ties) yields, as expected, estimates that often are closer to the model conditional on the degrees than to the model without vertex parameters. It is natural that for the subgraph counts, this model yields larger standard deviations than for the model conditioning on the degrees.

With respect to computing time, the model conditioning on all degrees was the quickest and had the best convergence properties. The restriction to this limited subset of the outcome space seems to facilitate the estimation algorithm.

A general conclusion is that these examples give some trust in *MCMC* estimation for exponential random graph models, conditional on the total number of edges. The *EIES* data example illustrates, however, that the variability (in this case: the relative standard deviations of the transitive triplets and the 3-cycles counts) in the estimated model still can be very large. Conditioning also on all degrees amounts to a model with a different interpretation; this model potentially 'explains away' part of the network structure, leads (in these examples) to larger standard errors, and to much smaller variability in the fitted networks. More research is needed to study when conditioning on the degrees is meaningful.

Appendix

A.1. Simulation-based estimation

The ML estimate is the solution of the moment equation (2). It is known from the general theory of exponential families that $\mu(\theta) = E_\theta u(Y)$ is the gradient of $\psi(\theta)$,

$$\mu_k(\theta) = \partial\psi(\theta)/\partial\theta_k; \tag{6}$$

that the covariance matrix

$$\Sigma(\theta) = (\sigma_{hk}(\theta))_{1\le h,k\le m} = \mathrm{cov}(u(Y))$$

of $u(Y)$ with elements $\sigma_{hk}(\theta)$ is the matrix of derivatives of $\mu(\theta)$,

$$\sigma_{hk}(\theta) = \frac{\partial\mu_k}{\partial\theta_h} = \frac{\partial^2\psi(\theta)}{\partial\theta_h\partial\theta_k}\ ; \tag{7}$$

and that the asymptotic covariance matrix of the ML estimator $\widehat{\theta}$ is given by

$$cov_\theta\left(\widehat{\theta}\right) = (\Sigma(\theta))^{-1}\ . \tag{8}$$

If $\mu(\theta)$ and $\Sigma(\theta)$ would be computable, the ML estimate could be found by the Newton-Raphson algorithm with iteration step

$$\widehat{\theta}^{(n+1)} = \widehat{\theta}^{(n)} - \left(\Sigma(\widehat{\theta}^{(n)})\right)^{-1}\left(\mu(\hat{\theta}^{(n)}) - u(y)\right)\ . \tag{9}$$

However, none of the functions $\psi(\theta)$, $\mu(\theta)$, or $\Sigma(\theta)$ can be computed in practice for exponential graph models, unless g is very small or the model is very simple (e.g., the reciprocity p^* model).

Snijders (2002) proposed to use the Robbins-Monro algorithm to solve the moment equation. This algorithm has iteration step

$$\widehat{\theta}^{(n+1)} = \widehat{\theta}^{(n)} - a_n D_n^{-1} Z(n)\ , \tag{10}$$

where $Z(n)$ is given by

$$Z(n) = u(Y^{(n)}) - u(y) \ ; \tag{11}$$

in this formula, $u(y)$ is the observed value of the sufficient statistic and $Y^{(n)}$ is a random draw from probability distribution (1) with parameter $\widehat{\theta}^{(n)}$. The step sizes a_n are a sequence of positive numbers converging to 0. Procedures for the Monte Carlo generation of $Y^{(n)}$, based on Gibbs sampling or the Metropolis-Hastings algorithm, are discussed in Snijders (2002). A constant positive diagonal matrix $D_n = D_0$ is used and a sequence a_n of the order $n^{-3/4}$. The estimate of θ is the average of the sequence $\widehat{\theta}^{(n)}$ or the average of the 'last part' of this sequence.

The algorithm proposed in Snijders (2002) consists of three 'phases'. The *first phase* is used to determine the diagonal matrix $D_n = D_0$ to be used in Phase 2 in the updating steps (10). The diagonal elements are estimates of the derivatives

$$d_{kk} = \partial E_\theta u_k(Y) = \partial \theta_k \ , \tag{12}$$

evaluated in the initial value θ of the estimation algorithm. The *second phase* iteratively determines provisional estimated values according to the updating steps (10). In the *third phase* the parameter value is kept constant at θ, the presumably found approximate solution of the moment equation (2). A large number of steps is carried out to check the approximate validity of this equation and estimate the estimation covariance matrix $\mathrm{cov}_\theta\big(\hat{\theta}\big)$.

A.2. Algorithm extension for vertex parameters

Including degrees and vertex parameters in the model leads to a high number of parameters. This is in itself not a big problem, as the algorithm includes no matrix inversions or other operations of order higher than $O(g)$. This section reviews the parts in the algorithm of Snijders (2002) that depend on the part of the model represented by (4).

Generating random (di)graphs according to the exponential random graph model can be carried out using MCMC algorithms based on the conditional probabilities

$$\ln\left(P_\theta\left\{Y_{ij} = y_{ij} \mid Y_{hk} = y_{hk} \text{ for all } (h,k) \neq (i,j)\right\}\right) \ , \tag{13}$$

or more involved conditional probabilities. The contribution to (13) corresponding to (4) is

$$\alpha_i y_{i+} + \beta_j y_{+j} \ . \tag{14}$$

Phase 1 of the algorithm is for the estimation of the diagonal elements (12). It follows from (6) and (7) that

$$d_{kk} = \partial E_\theta u_k(Y)/\partial\theta_k = \text{var } u_k(Y) \ . \tag{15}$$

In this case, the statistics $u_k(Y)$ are, respectively, the out-degrees Y_{i+} and the in-degrees Y_{+j} . For the values d_{kk} only very rough approximations are required. Therefore it is sufficient to have estimates of the variances (15) under the model where these degrees have binomial distributions. For a variable S with a binomial distribution with parameters $g-1$ and p, the estimated variance is $S(g-1-S)/(g-1)$. Therefore, for parameters with index k corresponding to the term $\alpha_i y_{i+}$ in the log-likelihood, d_{kk} can be defined by

$$d_{kk} = \frac{y_{i+}\left(g-1-y_{i+}\right)}{g-1} \ ,$$

and for parameters corresponding to the term $\beta_j y_{+j}$, by

$$d_{kk} = \frac{y_{+j}\left(g-1-y_{+j}\right)}{g-1} \ .$$

To keep these formulae positive for all values of the degrees, the degrees should first be truncated to the interval $[1, ..., g-2]$.

With these values, the updating steps for the parameters α_i are

$$\widehat{\alpha}_i^{(n+1)} = \widehat{\alpha}_i^{(n)} - a_n \frac{(g-1)(Y_{i+}^{(n)} - y_{i+})}{y_{i+}(g-1-y_{i+})} \ , \tag{16}$$

where the numerator $y_{i+}(g-1-y_{i+})$ must be truncated to a positive value. The updating steps for β_j are analogous.

A.3. Algorithm extension for conditioning on in- and out-degrees

The maximum likelihood estimation conditional on in- and out-degrees can be carried out by the MCMC method of Snijders (2002), provided that we have a way of generating exponential random graphs under this condition. An algorithm for the latter purpose can be constructed using the Metropolis Hastings algorithm with proposal steps that follow from Rao, Jana, and Bandyopadhyay (1996).

These authors show that it is possible to go from any digraph to any other digraph with the same in- and out-degrees by finitely many steps of the following kind. These steps are presented as changes for the adjacency matrix.

1. *Switching alternating rectangles.*

   For four distinct vertices i, j, h, k, with $y_{ih} = y_{jk} = 1$, $y_{ik} = y_{jh} = 0$, switch the values of these four elements (i.e., replace 0 by 1 and vice versa).

2. *Switching alternating triads.*

   (Rao et al. (1996) use the term compact alternating hexagons rather than alternating triads.) For three distinct vertices i, j, h with $y_{ij} = y_{jh} = y_{hi} = 1$, $y_{ji} = y_{hj} = y_{ih} = 0$, switch the values of these six elements.

The proposal distribution is a mixture, in proportions p and $1-p$, of switching a randomly selected alternating rectangle and switching triad. It was noted by Rao et al. (1996) that the number of possibilities for such switches depends on the current digraph. These authors proposed a way for dealing with this number of possibilities that requires to calculate the number of these possibilities. To circumvent this (somewhat time-consuming and error-prone) calculation, we propose a different and simpler technique.

For switching a rectangle, four distinct vertices i, j, h, k are selected at random. If $y_{ih} = y_{jk} \neq y_{ik} = y_{jh}$, this is an alternating rectangle. Denote by $\widetilde{y}$ the digraph that is obtained by switching this rectangle. The switch is carried out with probability

$$\min \left\{ 1, e^{\theta(u(\widetilde{y})-u(y))} \right\} .$$

If the condition $y_{ih} = y_{jk} \neq y_{ik} = y_{jh}$ is not satisfied, then the digraph remains unchanged at this iteration. The procedure for switching triads is entirely analogous. We use the proportions $p = 1 - p = 0.5$.

This proposal distribution has a constant probability for each switch of an alternating rectangle, and likewise for alternating triads. The probability of keeping the current digraph unchanged is variable, but this probability does not occur in the formula or convergence proof of the Metropolis Hastings procedure.

References

[1] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society,* ser. B, 36, 192 - 225.

[2] Corander, J., Dahmström, K. and Dahmström, P. (1998). *Maximum likelihood estimation for Markov graphs.* Research Report 1998:8, Department of Statistics, Stockholm University.

[3] Dahmström, K., and Dahmström, P. (1993). *ML-estimation of the clustering parameter in a Markov graph model.* Research report, Department of Statistics, Stockholm University.

[4] Frank, O. (1991). Statistical analysis of change in networks. *Statistica Neerlandica*, 45, 283 - 293.

[5] Frank, O., and D. Strauss. (1986). Markov graphs. *Journal of the American Statistical Association*, 81, 832 - 842.

[6] Freeman, L.C., and Freeman, S.C. (1980). A semi-visible college: Structural effects on a social networks group. Pp. 77 - 85 in Henderson, M.M., and McNaughton, M.J. (eds). *Electronic Communication: Technology and Impacts.* Boulder, CO: Westview Press.

[7] Geyer, C.J., and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society*, ser. B, 54, 657 - 699.

[8] Holland, P.W. and Leinhardt, S. (1975). Local structure in social networks. In D. Heise (ed)., *Sociological Methodology - 1976*. San Francisco: Jossey-Bass.

[9] Holland, P.W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76, 33 - 65.

[10] Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9, 109 - 134.

[11] Lehmann, E.L. (1983). *Theory of Point Estimation*. New York: Wiley.

[12] Pattison, P., andWasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52, 169 - 193.

[13] Rao, A.R., Jana, R. and Bandyopadhyay, S. (1996). A Markov chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhyã*, series A, 58, 225-242.

[14] Robbins, H., and Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, 22, 400 - 407.

[15] Robins, G., Pattison, P. and Wasserman, S. (1999). Logit models and logistic regressions for social networks, III. Valued relations. *Psychometrika*, 64, 371 - 394.

[16] Snijders, T.A.B. (1981). The degree variance: an index of graph heterogeneity. *Social Networks*, 3, 163 - 174.

[17] Snijders, T.A.B. (1991). Enumeration and simulation methods for 0-1 matrices with given marginals. *Psychometrika*, 56, 397 - 417.

[18] Snijders, T.A.B. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*, 21, 149 - 172. Also published in Doreian and Stokman(1997).

[19] Snijders, T.A.B. (2001). The Statistical Evaluation of Social Network Dynamics. M.E. Sobel and M.P. Becker (eds)., *Sociological Methodology - 2001*, 361 - 395. Boston and London: Basil Blackwell.

[20] Snijders, T.A.B. (2002). Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure* vol. 3 no. 2.

[21] Snijders, T.A.B., and Huisman, M. (2002). *Manual for* SIENA *version 1.96.*Groningen: ICS.

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

[23] Wasserman, S. (1977). Random directed graph distributions and the triad census in social networks. *Journal of Mathematical Sociology*, 5, 61 - 86.

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

[25] Wasserman, S., and Pattison, P. (1996). Logit models and logistic regression for social networks: I. An introduction to Markov graphs and $p*$. *Psychometrika*, 61, 401 - 425.