2014

Delayed Rejection Algorithm to Estimate Bayesian Social Networks

Alberto Caimo
Technological University Dublin, alberto.caimo@tudublin.ie

Antonietta Mira
Università della Svizzera Italiana, antonietta.mira@usi.ch

Recommended Citation

Delayed rejection algorithm to estimate
Bayesian social networks

Alberto Caimo

InterDisciplinary Institute of Data Science, University of Lugano, Switzerland
E-mail: alberto.caimo@usi.ch

Antonietta Mira

InterDisciplinary Institute of Data Science, University of Lugano, Switzerland
E-mail: antonietta.mira@usi.ch

Summary: Statistical social network analysis has become a very active and fertile area of research in the recent past. Recent developments in Bayesian computational methods have been successfully applied to estimate social network models. The Delayed rejection (DR) strategy is a modification of the Metropolis-Hastings (MH) algorithms that reduces the variance of the resulting Markov chain Monte Carlo estimators and allows partial adaptation of the proposal distribution. In this paper we show how the DR strategy can be exploited to estimate dyadic independence social network models leading to an average 40% variance reduction relative to the competing MH algorithm, confirming that DR dominates, in terms of Peskun ordering, the MH algorithm.

Keywords: Delayed rejection; Peskun ordering; social networks; Bayesian analysis.

1. Dyadic independence network models

Statistical network models (see Salter-Townshend et al. (2012) for a recent review) are widely used in many scientific areas as they give the possibility to investigate how an

observed network may be related to local relational structures. Social network analysis is based on the study of social relations between actors so as to understand the creation of global relational structure from the basic local relation. From a statistical viewpoint, social networks are relational data represented as a graph consisting of a set of n nodes and a set of m edges which define some sort of relations between pairs of nodes called dyads. The network structure is described by an $n \times n$ adjacency matrix $\boldsymbol{y}$ with elements:

$$y_{ij} = \begin{cases} 1, & \text{if node } i \text{ and node } j \text{ are connected} \\ 0, & \text{otherwise.} \end{cases}$$

If $y_{ij} = y_{ji}$ then the adjacency matrix is symmetric and the graph is undirected, otherwise the graph is directed and it is often called digraph.

The interest in network modeling began with a series of papers by Gilbert (1959) and Erdös and Rényi (1961) who set up the definition of *random graph* under the basic assumption that the presence of an edge between two nodes occurs with a fixed probability θ:

$$p(\boldsymbol{y}|\theta) = \prod_{i \neq j} \theta^{y_{ij}} (1 - \theta)^{1 - y_{ij}}$$

The likelihood of the model is a binomial distribution and implies a uniform distribution over the sample space of possible graphs with exactly m edges. The expected probability p of every edge is $m/\binom{n}{2}$ and implies the expected number of edges $p\binom{n}{2}$. In essence, this means that the model imposes that every node in a graph has approximately the same number of neighbours. The simplicity of this assumption makes the model tractable but inadequate to describe realistic phenomena as there are very few real-world networks with such simple structure. The need of formal tools for assessing what kinds of models are more appropriate has been at the basis of the development of models which led to the definition of the exponential random graph models.

The *beta model*, recently defined by Blitzstein and Diaconis (2010) and Chatterjee et al. (2011), is the natural heterogenous version of the random graph model and satisfies the dyadic independence property according to which dyads are assumed to be statistically independent. The model assumes that the nodal degree sequence captures the information in a network so that different graphs with the same degree sequence are considered equally likely.

The p_1 *model* proposed by Holland and Leinhardt (1981) can be considered the directed version of the beta model and expresses the presence of three tendencies: the

propensity with which a node i will be connected to j; the propensity with which a node i will attract others; and the degree of reciprocated edges. The p_2 *model* (van Duijn et al., 2004) is a random effects version of the p_1 model which recognises dependence between dyads with same nodes.

All the models described above belong to exponential random graph models (Lusher et al., 2012) that assume that the topological structure in an observed network $\boldsymbol{y}$ can be explained by the relative prevalence of a set of overlapping sub-graph configurations $s(\boldsymbol{y})$ also called graph or network statistics. Each network statistic is assumed to have a particular probability of being observed in the given network: higher is the probability of being expressed in the graph, more are the chances of that statistic to occur and vice versa. The probability of a network statistics being present a network is expressed in terms of parameters $\boldsymbol{\theta}$. Statistics with a positive parameter value for $\boldsymbol{\theta}$ have a greater than chance probability of being observed in any graph represented by the model and vice versa. ERGMs can be expressed mathematically as:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^t s(\boldsymbol{y})\}}{z(\boldsymbol{\theta})} \tag{1}$$

where $z(\boldsymbol{\theta})$ is a normalising constant.

2. MCMC methods for Bayesian social networks

Bayesian inference aims at learning about model parameters $\boldsymbol{\theta}$ that characterize the data generating process given the observed data $\boldsymbol{y}$ allowing one to explicitly incorporate prior beliefs and knowledge. In fact, in a Bayesian perspective the unknown parameters are considered as random variables and therefore are treated probabilistically.

The quantity of interest is the posterior distribution which computed using the Bayes' Theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{y})} \tag{2}$$

The posterior represents the conditional distribution of the parameters given the observed data while the prior $p(\boldsymbol{\theta})$ represents the distribution of the parameters before having observed the data. Therefore the posterior allows one to make probabilistic statements about how likely parameter values are after observing the data.

The prior distribution contains all the a priori information about the parameters while the posterior distribution has the function of translating the likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$ into a probability distribution that can be summarised as any probability distribution by computing expected values, standard deviations, quantiles, etc. Summarising the posterior distribution analytically is often impossible due to the intractability of the prior predictive distribution of the observed data:

$$p(\boldsymbol{y}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta}) \, p(\boldsymbol{\theta}) \, d\boldsymbol{\theta}. \tag{3}$$

Over the last twenty years, Monte Carlo methods have been used by Bayesian statisticians to perform a numerical approximation of the posterior distribution.

Bayesian methods are becoming increasingly popular as techniques for modelling social networks (Koskinen et al., (2010); Caimo and Friel, 2011; Caimo and Friel, 2013; Caimo and Mira, 2015). Direct evaluation of $p(\boldsymbol{\theta}|\boldsymbol{y})$ requires the calculation of both the likelihood $p(\boldsymbol{y}|\boldsymbol{\theta})$, which is computationally demanding if not intractable, and the marginal likelihood $p(\boldsymbol{y})$ which is typically intractable.

2.1. Delayed rejection strategy

Markov chain Monte Carlo (MCMC) algorithms are general simulation methods for sampling from posterior distributions and computing posterior quantities of interest.

Delayed rejection (DR) is a variation of the Metropolis-Hastings (MH) algorithm (Tierney, 1994; Green and Mira, 2001; Mira, 2001a) aimed at improving efficiency of the resulting MCMC estimators relative to Peskun (1973); Tierney (1998); and Mira (2001b) asymptotic variance ordering.

The key idea behind the algorithm is that, upon rejection in a MH, instead of advancing time and retaining the same position, a second stage move is proposed. This way it is possible to improve the MH algorithm by reducing the number of rejected candidates. The acceptance probability of the second stage candidate preserves reversibility of the Markov chain with respect to the target posterior density. This delaying rejection mechanism can be iterated for a fixed or random number of stages. The higher stage proposal distributions are allowed to depend on the candidates so far proposed and rejected. Thus DR allows partial local adaptation of the proposal within each time step of the Markov chain still retaining reversibility and the Markovian property.

The advantage of DR over alternative ways of combining different MH proposals or kernels such as mixing and cycling (Tierney, 1994), is that a hierarchy between kernels can be exploited so that the kernels that computationally intensive kernels are tried first. Or moves that are more "bold" (bigger variance of the proposal, for example) are tried at earlier stages thus allowing the sampler to explore the state space more efficiently following a sort of 'first bold' versus 'second timid' tennis-service strategy.

To simplify the notation indicate the posterior distribution of interest, $p(\boldsymbol{\theta}|\boldsymbol{y})$, simply as $p(\boldsymbol{\theta})$, dropping the conditioning on y. Suppose the current state of the Markov chain is $X_t = \boldsymbol{\theta}$. As in a standard MH, a candidate move $\boldsymbol{\theta}_1$ is generated from a proposal $q_1(\boldsymbol{\theta}, \cdot)$ and accepted with probability

$$
\begin{aligned}
\alpha_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1) &= 1 \wedge \frac{p(\boldsymbol{\theta}_1)q_1(\boldsymbol{\theta}_1, \boldsymbol{\theta})}{p(\boldsymbol{\theta})q_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)} \\
&= 1 \wedge \frac{N_1}{D_1}.
\end{aligned}
$$

Whenever a candidate $\boldsymbol{\theta}_1$ is rejected, instead of retaining the current state of a Markov chain as its new state $X_{t+1} = \boldsymbol{\theta}$, a second stage move $\boldsymbol{\theta}_2$ is generated from a proposal distribution that is allowed to depend, not only on the current position of the chain $\boldsymbol{\theta}$, but also on $\boldsymbol{\theta}_1$: $q_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \cdot)$. The second stage acceptance probability can be written as:

$$
\begin{aligned}
\alpha_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= 1 \wedge \frac{p(\boldsymbol{\theta}_2)q_1(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1)q_2(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta})[1 - \alpha_1(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1)]}{p(\boldsymbol{\theta})q_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)q_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)[1 - \alpha_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)]} \\
&= 1 \wedge \frac{N_2}{D_2}.
\end{aligned}
$$

This process of delaying rejection can be iterated (Mira 2001a) and the i-th stage acceptance probability is:

$$
\begin{aligned}
\alpha_i(\boldsymbol{\theta}, \boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_i) &= 1 \wedge \frac{N_i}{D_i} \\
&= 1 \wedge \left\{ \frac{p(\boldsymbol{\theta}_i)q_1(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{i-1})q_2(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i-2}) \cdots q_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{i-1} \cdots \boldsymbol{\theta})}{p(\boldsymbol{\theta})q_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)q_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \cdots q_i(\boldsymbol{\theta}, \boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_i)} \right. \\
&\quad \left. \frac{[1 - \alpha_1(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{i-1})][1 - \alpha_2(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i-2})] \cdots [1 - \alpha_{i-1}(\boldsymbol{\theta}_i, \cdots, \boldsymbol{\theta}_1)]}{[1 - \alpha_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)][1 - \alpha_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)] \cdots [1 - \alpha_{i-1}(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{i-1})]} \right\}
\end{aligned}
$$

If i-th stage is reached, it means that $N_j < D_j$ for $j = 1, \cdots, i-1$, therefore $\alpha_j(\boldsymbol{\theta}, \boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_j)$ is simply N_j/D_j, $j = 1, \cdots, i-1$ and a recursive formula can be obtained: $D_i = q_i(\boldsymbol{\theta} \cdots \boldsymbol{\theta}_i)(D_{i-1} - N_{i-1})$ which leads to:

$$D_i = q_i(\boldsymbol{\theta} \cdots \boldsymbol{\theta}_i)[q_{i-1}(\boldsymbol{\theta} \cdots \boldsymbol{\theta}_{i-1})[q_{i-2}(\boldsymbol{\theta}, \cdots, \boldsymbol{\theta}_{i-2}) \cdots$$
$$[q_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)[q_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)p(\boldsymbol{\theta}) - N_1] - N_2] - N_3] \cdots - N_{i-1}].$$

Since reversibility with respect to p is preserved separately at each stage, the process of delaying rejection can be interrupted at any stage. It is therefore possible to decide, in advance, to try at most, a fixed number of moves away from the current state or, alternatively, upon each rejection, move to a higher stage proposal with probability p or otherwise stay in the current state. The DR strategy provides MCMC estimators with smaller asymptotic variance than standard MH (Tierney and Mira, 1999).

3. Example

One of the most popular statistical network model for large network datasets is the beta model. This can be formalised by saying that it is an exponential random graph model with the degree sequence as a sufficient statistic vector.

If $\boldsymbol{y}$ is the observed network graph with degree sequence $d_1, \ldots, d_n$ the likelihood of the beta model can be written as:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{\exp\left\{\sum_i^n \theta^i \, d_i\right\}}{\prod_{i<j}(1 + \exp\left\{\theta^i + \theta^j\right\})}$$

where θ^i indicates the i-th parameter in the $\boldsymbol{\theta}$ vector.

The Zachary's karate club network dataset (Figure 1, Zachary, 1977) represents the social network of friendships between 34 members of a karate club at a US university in the 1970.

We will use this data to compare MH and DR in terms of efficiency. Each iteration of the MH algorithm consists in generating a new vector of parameters $\boldsymbol{\theta}_1$ from some proposal distribution $q_1(\cdot)$. A block-update sampler with normal proposal is used to simultaneously update all the parameter of the posterior distribution.

The DR algorithm that we propose here consists of two stages. The first stage proposal distribution is the same as the proposal used in the competing MH sampler and the first stage acceptance probability is:
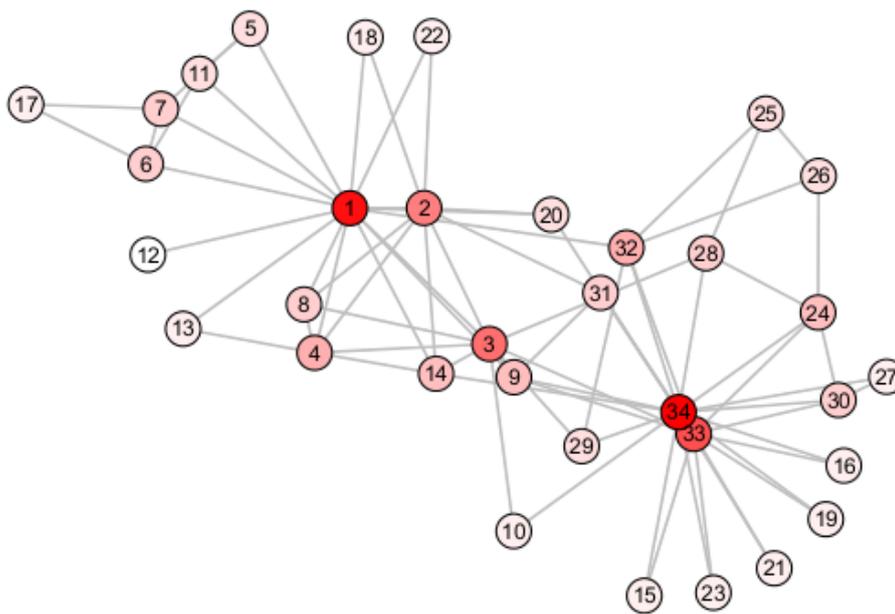
Figure 1. Zachary karate club network graph: social network of friendships between 34 members of a karate club at a US university in the 1970. The more red the color of a node, the greater its degree.

$$\alpha_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1) = 1 \wedge \frac{p(\boldsymbol{y}|\boldsymbol{\theta}_1)p(\boldsymbol{\theta}_1)q_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)}{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})q_1(\boldsymbol{\theta}_1, \boldsymbol{\theta})}.$$

The second stage of the DR algorithm, consists in generating $\boldsymbol{\theta}_2$ from a different proposal distribution $q_2(\cdot)$. The second stage acceptance probability is defined as:

$$\alpha_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1, \theta_2) = 1 \wedge \frac{p(\boldsymbol{y}|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2)q_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)q_2(\boldsymbol{\theta}, \boldsymbol{\theta}_1)[1 - \alpha_1(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1)]}{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})q_1(\boldsymbol{\theta}_1, \boldsymbol{\theta})q_2(\boldsymbol{\theta}_2, \boldsymbol{\theta})[1 - \alpha_1(\boldsymbol{\theta}, \boldsymbol{\theta}_1)]}.$$

The posterior estimates displayed in Table 1 were obtained using very flat normal prior for each parameter: $p(\theta^i) \sim \mathcal{N}(0, \sigma^2)$ where $\sigma^2 = 100$, a normal proposal distribution $q_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\theta}, \sigma_{h1}^2 \boldsymbol{I}_n)$ where $\sigma_{h1}^2 = 0.06$ (in order to reach an acceptance rate of about 21%) and $\boldsymbol{I}_n$ is the n-dimensional identity matrix, and $10,000$ MCMC iterations.

The second stage proposal is as a deterministic move in the opposite direction of the move at first stage so that: $\boldsymbol{\theta}_2 = 2\boldsymbol{\theta} - \boldsymbol{\theta}_1$. We consider, as function of interest, the identity function, i.e. $f(\theta^i) = \theta^i, \forall i$ in other words, we focus on estimating the posterior mean of each parameter of interest.

We report (Table 1) the effective sample sizes (Kass et al., 1998) of the DR (ESS_{DR}) and MH (ESS_{MH}) for each parameter of the model. The average ESS for the DR and MH are about 157 and 74 respectively. The CPU simulation times obtained by the DR and MH are respectively: $TIME_{DR} = 69$ secs and $TIME_{MH} = 47$ secs meaning the MH is faster but less efficient than the DR.

The efficiency ratios can be calculated from the ESS and CPU simulation times:

$$\frac{ESS_{DR}/ESS_{MH}}{TIME_{DR}/TIME_{MH}}.$$

In this example, the efficient ratio is about 1.4 meaning that the DR procedure is 40% more efficient than the MH algorithm. The analysis has been conducted with the Bergm package for R (Caimo and Friel, 2014).

There are many other possible second stage proposal strategies that can be proposed. In this example, if we vary the proposal variance σ_{h1}^2 we can get different performance results. Using this second stage proposal strategy, a high value for σ_{h1}^2 implies a 'bold' move at the first stage and therefore a low first stage acceptance rate. This may result in a decreasing performance of the DR algorithm because the second stage proposal defined

above involves the same bold move in the opposite direction. In this case, a 'timid' second stage proposal, for example, would work better.

The presence of effects associated to nodal degrees has an important impact in explaining the overall network structure. In fact, positive parameter values correspond to nodes with high degree and negative parameter values correspond to nodes with low degree. The results of Table 1 are consistent with previous results demonstrating the social prominence of the actors associated to node 1 and 34. Zachary (1977) studied conflict and fission in this network, as the karate club was split into two separate clubs, after long disputes between two factions of the club, one led by John A. (node 34 in Figure 1), the other by Mr. Hi (node 1 in Figure 1). The use of the beta model in this context is useful as it allow us to understand how the relational structure is influenced by the popularity of the two competing leaders and the behaviour of the other actors that eventually led to the split into seperate clubs.

3.1. Goodness of fit diagnostics

Estimation procedures need to be tested for their accuracy through network simulation based on the estimated parameter values and comparison of simulated graphs with the observed graph through the calculation of goodness of fit statistics. The sample graphs are analysed as a distribution of possible graphs. By comparing the network statistics of the original observed graph with the graph statistics of the simulated networks it is possible to assess the goodness of fit of the social networks (Hunter et al., 2008). This set of GOF statistics generally includes many measurable network statistics contained in the observed and simulated graphs. This means that configurations that are not included in the ERGM are actually still tested to see if they are similar in the observed and simulated graphs.

In Figure 2 we can se that, based on various GoF statistics, the networks simulated from the estimated posterior distribution are in reasonable agreement the observed network. We can therefore conclude that the model is a reasonable fit to the data.
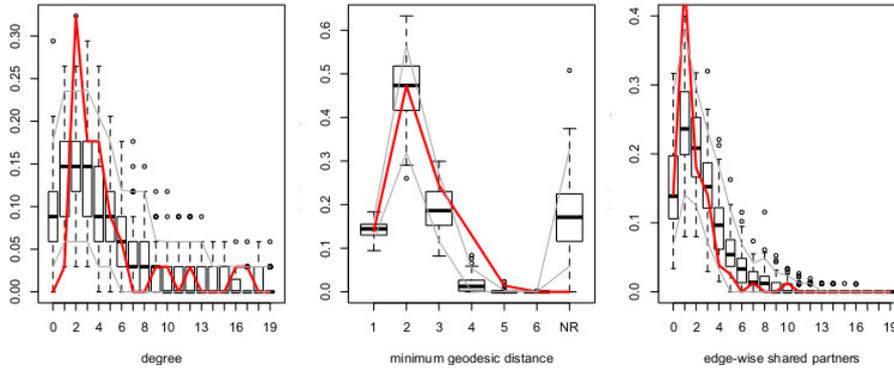
Figure 2. GoF diagnostics: The solid red line displays the goodness of fit statistics for the observed data together with boxplots of GoF network statistics based on 100 simulated networks from the posterior distribution.

4. Conclusions

The better performance of the delayed rejection strategy relative to the standard MH sampler is demonstrated in the setting of dyadic independence Bayesian social network models using the well known Zachary karate club network. A very simple deterministic second stage proposal is considered for the delayed rejection algorithm but more creative strategies could be elaborated. Performance comparison takes CPU simulation time into account and considers the effective sample size of the two competing samplers. For all the 34 parameters of the beta model, the DR algorithm leads to a decrease of autocorrelation along the path of the simulated Markov chain which, in turn, translates into a smaller asymptotic variance and a higher effective sample size. Averaging over the different parameters, the efficiency ratio of the DR is higher than the one of the corresponding MH by approximately 40%.

References

Blitzstein, J. and Diaconis, P. (2010), A sequential importance sampling algorithm for generating random graphs with prescribed degrees, *Internet Mathematics*, **6**(4), 489–522.

Caimo, A. and Friel, N. (2011), Bayesian Inference for Exponential Random Graph Models, *Social Networks*, **33**, 41–55.

Caimo, A. and Friel, N. (2013), Bayesian model selection for exponential random graph models, *Social Networks*, **35**, 11–24.

Caimo, A. and Friel, N. (2014), Bergm: Bayesian Exponential Random Graphs in R, *Journal of Statistical Software*, **61**(2), 1–25.

Caimo, A. and Mira, A. (2015), Efficient Computational Strategies for Doubly Intractable Problems with Applications to Bayesian Social Networks, *Statistics and Computing*, **25**(1), 113–125.

Chatterjee, S., Diaconis, P., and Sly, A. (2011), Random graphs with a given degree sequence, *Annals of Applied Probability*, **21**(4), 1400–1435.

Erdös, P. and Rényi, A. (1961), On the strength of connectedness of a random graph, *Acta Mathematica Scientia Hungary*, **12**, 261–267.

Gilbert, E. N. (1959), Random graphs, *Annals of Mathematical Statistics*, **30**, 1141–1144.

Green, P. J. and Mira, A. (2001), Delayed rejection in reversible jump Metropolis–Hastings, *Biometrika*, **88**, 1035–1053.

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.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008), Goodness of fit of social network models, *Journal of the American Statistical Association,* **103**, 248–258.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998), Markov chain monte carlo in practice: A roundtable discussion, *The American Statistician*, **52**, 93–100.

Koskinen, J. H., Robins, G. L., and Pattison, P. E. (2010). Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation, *Statistical Methodology*, **7**(3), 366–384.

Lusher, D., Koskinen J., and Robins G. (2012), Exponential random graph models

for social networks: Theory, methods, and applications. Cambridge University Press.

Mira, A. (2001a), On Metropolis-Hastings algorithms with delayed rejection, *Metron*, **59**, 231–241.

Mira, A. (2001b), Ordering and improving the performance of Monte Carlo Markov chains, *Statistical Science*, 340–350.

Peskun, P. (1973), Optimum monte-carlo sampling using markov chains, *Biometrika*, **60**, 607–612.

Salter-Townshend, M., White, A., Gollini, I., and Murphy, T. B. (2012), Review of statistical network analysis: models, algorithms, and software, *Statistical Analysis and Data Mining*, **5**, 243–264.

Tierney, L. (1994), Markov chains for exploring posterior distributions, *the Annals of Statistics*, 1701–1728.

Tierney, L. (1998), A note on Metropolis-Hastings kernels for general state spaces, *Annals of Applied Probability*, 1–9.

Tierney, L. and Mira, A. (1999), Some adaptive Monte Carlo methods for Bayesian inference, *Statistics in medicine*, **18**, 2507–2515.

van Duijn, M. A., Snijders, T. A. B., and Zijlstra, B. H. (2004), p2: a random effects model with covariates for directed graphs, *Statistica Neerlandica*, **58**, 234–254.

Zachary, W. (1977), An Information Flow Model for Conflict and Fission in Small Groups, *Journal of Anthropological Research*, **33**, 452–473.

Table 1. *Posterior mean (PM), posterior standard deviation (PSD) and efficiency ratio (ER) estimates for each parameter obtained by the DR algorithm. Effective sample sizes (ESS) for both DR and MH.*

	θ^1	θ^2	θ^3	θ^4	θ^5	θ^6	θ^7	θ^8	θ^9	θ^{10}
PM	1.42	0.13	0.37	-0.63	-1.71	-1.28	-1.28	-1.31	-0.89	-2.27
PSD	0.42	0.47	0.45	0.53	0.67	0.60	0.65	0.62	0.59	0.81
ESS_{MH}	187	113	166	112	60	56	78	79	98	32
ESS_{DR}	319	292	235	219	125	195	175	166	184	104
	θ^{11}	θ^{12}	θ^{13}	θ^{14}	θ^{15}	θ^{16}	θ^{17}	θ^{18}	θ^{19}	θ^{20}
PM	-1.78	-3.29	-2.44	-0.93	-2.35	-2.33	-2.37	-2.31	-2.31	-1.75
PSD	0.74	1.11	0.88	0.57	0.81	0.82	0.87	0.82	0.80	0.70
ESS_{MH}	61	10	36	88	28	42	47	48	44	57
ESS_{DR}	129	46	96	202	100	96	100	96	111	121
	θ^{21}	θ^{22}	θ^{23}	θ^{24}	θ^{25}	θ^{26}	θ^{27}	θ^{28}	θ^{29}	θ^{30}
PM	-2.28	-2.33	-2.39	-0.91	-1.81	-1.72	-2.31	-1.31	-1.77	-1.30
PSD	0.81	0.83	0.89	0.57	0.69	0.72	0.81	0.64	0.69	0.62
ESS_{MH}	57	20	18	89	73	89	35	68	70	65
ESS_{DR}	87	103	98	158	111	117	110	149	124	161
	θ^{31}	θ^{32}	θ^{33}	θ^{34}						
PM	-1.31	-0.61	0.74	1.57						
PSD	0.62	0.52	0.43	0.41						
ESS_{DR}	88	101	173	142						
ESS_{MH}	155	205	324	334						