



Statistical Network Analysis with Bergm

Alberto Caimo 
TU Dublin

Lampros Bouranis 
AUEB Athens

Robert Krause
University of
Kentucky

Nial Friel 
UCD Dublin

Abstract

Recent advances in computational methods for intractable models have made network data increasingly amenable to statistical analysis. Exponential random graph models (ERGMs) emerged as one of the main families of models capable of capturing the complex dependence structure of network data in a wide range of applied contexts. The **Bergm** package for R has become a popular package to carry out Bayesian parameter inference, missing data imputation, model selection and goodness-of-fit diagnostics for ERGMs. Over the last few years, the package has been considerably improved in terms of efficiency by adopting some of the state-of-the-art Bayesian computational methods for doubly-intractable distributions. Recently, version 5 of the package has been made available on CRAN having undergone a substantial makeover, which has made it more accessible and easy to use for practitioners. New functions include data augmentation procedures based on the approximate exchange algorithm for dealing with missing data, adjusted pseudo-likelihood and pseudo-posterior procedures, which allow for fast approximate inference of the ERGM parameter posterior and model evidence for networks on several thousands nodes.

Keywords: Bayesian inference, exponential random graph models, R packages.

1. Introduction

Exponential random graph models (ERGMs, Frank and Strauss 1986; Wasserman and Pattison 1996; Robins, Pattison, Kalish, and Lusher 2007) are one of the most important families of statistical models conceived to capture the complex dependence structure of an observed network and identify the relational effects that are supposed to describe the link creation process.

Bayesian inference for ERGMs is challenging because of the intractability of both the likelihood and the marginal likelihood. The advanced computational methods developed by several

recent papers (see for example, Koskinen, Robins, and Pattison 2010; Caimo and Friel 2011; Caimo and Mira 2015; Alquier, Friel, Everitt, and Boland 2016; Bouranis, Friel, and Maire 2017, 2018) have made it possible and computationally feasible to model increasingly large network data using ERGMs on several thousands of nodes. The development of user-friendly software has always represented an essential aspect of the research activity in this area.

There is a wide range of R packages (R Core Team 2022) implementing various inferential approaches and modeling extensions of the ERGM framework. These include the **ergm** package (Hunter, Handcock, Butts, Goodreau, and Morris 2008b) providing a comprehensive set of functions for fitting, simulating and diagnosing ERGMs; the **tergm** (Krivitsky and Handcock 2014) and **btergm** (Leifeld, Cranmer, and Desmarais 2018) packages for the analysis of temporal ERGMs; the **mlergm** (Stewart and Schweinberger 2018) package for analyzing multilevel ERGMs; and the **fergm** package (Box-Steffensmeier, Christenson, and Morgan 2018) implementing estimation and fit assessment for frailty ERGMs. The main software alternative to these R packages is the free **PNet** program (Wang, Robins, and Pattison 2009) which runs on Windows.

The **Bergm** package for R implements Bayesian analysis for exponential random graph models, providing a comprehensive inferential framework for Bayesian parameter estimation and model selection using efficient Monte Carlo algorithms. It can also supply model assessment and goodness-of-fit procedures that address the issue of model adequacy. Although computationally intensive, the package is easy to use and represents an attractive way of analyzing networks by adopting a fully probabilistic treatment of uncertainty of the network effects that are assumed to be able to explain the overall connectivity structure of the network. The **Bergm** package depends on the **ergm** package which is part of the **statnet** suite of packages (Handcock, Hunter, Butts, Goodreau, and Morris 2008) and therefore it makes use of the same model set-up and network simulation algorithms. The **Bergm** package has been improved considerably in terms of usability for practitioners and performance since its early versions (Caimo and Friel 2014). The aim of the package is to provide a set of tools for both developers and end-users. For this reason the package includes several competing functions (based on different statistical approaches) for accomplishing posterior parameter estimation and model selection.

The package has been used in several applications such as neuroscience (Sinke, Dijkhuizen, Caimo, Stam, and Otte 2016), organization science (Caimo and Lomi 2014; Tasselli and Caimo 2019) and political science (Henning, Aßmann, Hedtrich, Ehrenfels, and Krampe 2019).

The purpose of this paper is to provide a complete description of the recent improvements by summarizing some of the technical background and newly implemented algorithms in the main functions of the package.

2. Getting Bergm

The **Bergm** package (Caimo, Bouranis, and Krause 2022) is available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=Bergm> and can be obtained in R using the following commands:

```
R> install.packages("Bergm")  
R> library("Bergm")
```

Bergm depends on **ergm** (Hunter *et al.* 2008b), **network** (Butts 2008), **coda** (Plummer, Best, Cowles, and Vines 2006), **MCMCpack** (Martin, Quinn, and Park 2011), **Matrix** (Bates and Mächler 2022), **mvtnorm** (Genz *et al.* 2021), **matrixcalc** (Novomestky 2012) and **Rglpk** (Theussl and Hornik 2019). Loading the package will automatically load all the dependencies. All of these packages are available on CRAN at <http://CRAN.R-project.org/>. The results presented in this paper have been obtained using R version 4.2.1; **Bergm** version 5.0.5; **ergm** version 4.2.2; **network** version 1.17.2; **coda** version 0.19-4; **MCMCpack** version 1.6-3; **Matrix** version 1.4-1; **mvtnorm** version 1.1-3; **matrixcalc** version 1.0-5 and **Rglpk** version 0.6-4.

3. Network data

Two very well known network datasets are used throughout this tutorial for illustrative purposes: the first is the Lazega’s Law Firm dataset whose undirected edges represent collaborative relations in a Northeastern US corporate law firm (Lazega 2001); the second is the Faux Dixon High School dataset, which represents a directed friendship network (Resnick *et al.* 1997).

3.1. Lazega’s law firm

The law office in Lazega’s study network (graphs displayed in Figure 1), consists of binary undirected collaborative relations between 36 partners in a Northeastern US corporate law firm (Lazega 2001). Various members’ attributes are also part of the dataset, including seniority, formal status, office in which they work, gender, law school attended, individual performance measurements (hours worked, fees brought in), attitudes concerning various management policy options.

```
R> data("lazega", package = "Bergm")
```

3.2. Faux Dixon high school

The Faux Dixon High School network data (graphs displayed in Figure 2), included in the **ergm** package, represents a simulation of a binary directed in-school friendship network (Resnick *et al.* 1997). See `?faux.dixon.high` for the ERGM that was fit to the original

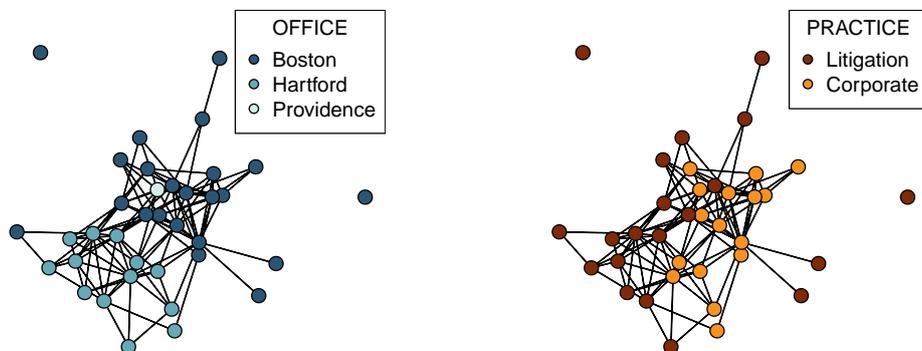


Figure 1: Lazega’s network undirected graphs.

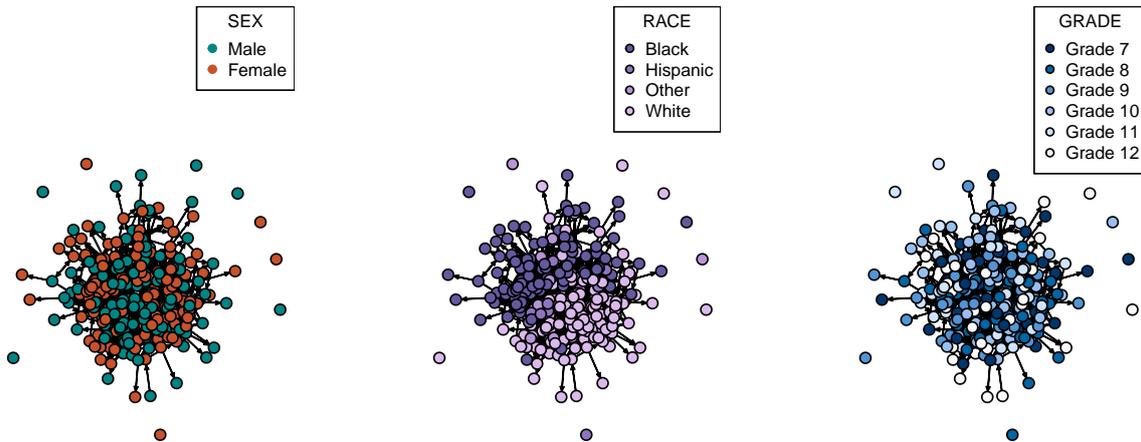


Figure 2: Faux Dixon High School network directed graphs.

data, generating the network dataset. The network comprises 248 nodes representing students. Information on the following nodal attribute variables is available: sex, race, grade.

```
R> data("faux.dixon.high", package = "ergm")
R> dixon <- faux.dixon.high
```

The code used to plot the network graphs in Figures 1 and 2 is included in the R replication code of the supplementary files.

4. Exponential random graph models

Networks are relational data defined as a collection of actors interacting with each other and connected in a pairwise fashion. Networks can be represented as graphs consisting of a set of n nodes and a set of edges which define some sort of relationships between pairs of nodes (dyads). The connectivity pattern of a graph can be described by an $n \times n$ adjacency matrix Y encoding the presence or absence of an edge between nodes i and j :

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

If the network is undirected, then $y_{ij} = y_{ji}$ and the adjacency matrix is symmetric, otherwise the network is directed. Edges connecting a node to itself (self-loops) are not allowed in many applications and will not be considered in this context.

Introduced by [Holland and Leinhardt \(1981\)](#) to model individual heterogeneity of nodes and reciprocity of their edges, the family of ERGMs was generalized by [Frank and Strauss \(1986\)](#), [Wasserman and Pattison \(1996\)](#) and [Snijders, Pattison, Robins, and Handcock \(2006\)](#) in order to account for higher-order dyadic relationships. ERGMs constitute a broad class of network models (see [Robins *et al.* 2007](#) for an introduction) assuming that the probability of an observed network y can be explained in terms of the relative prevalence of a set of network

statistics $s(y)$:

$$f(y | \theta) = \frac{\exp\{\theta^\top s(y)\}}{z(\theta)}, \quad (1)$$

where $\theta \in \Theta \subseteq \mathbb{R}^d$ is the vector of d model parameters associated with $s(y)$ (Hunter and Handcock 2006; Snijders *et al.* 2006), $z(\theta) = \sum_{y \in \mathcal{Y}} \exp\{\theta^\top s(y)\}$ is the likelihood normalizing constant consisting of a sum across \mathcal{Y} which is the set of all possible graphs on n nodes. Additionally, ERGMs allow for incorporation of covariate information X , e.g., the number of network statistic configurations within the same attribute category.

The normalizing constant $z(\theta)$ is computationally tractable only for trivially-small networks, as \mathcal{Y} involves $2^{\binom{n}{2}}$ and $2^{n(n-1)}$ possible undirected and directed graph configurations respectively. The network statistics $s(y)$ often represent a series of counts of sub-graph configurations (e.g., the number of edges, stars, triangles, functions of degree distributions, edgewise shared partners), that capture the relevant information of the global connectivity structure of the network graph (Snijders *et al.* 2006).

Given Equation 1 we can express the distribution of the Bernoulli variable Y_{ij} under the conditional form

$$\text{logit}\{\mathbb{P}(Y_{ij} = 1 | y_{-ij}, \theta)\} = \theta^\top \delta_s(y)_{ij},$$

where $\delta_s(y)_{ij} = s(y_{ij}^+) - s(y_{ij}^-)$ denotes the vector representing the change in the vector of network sufficient statistics when the value of y_{ij} is toggled from a 0 (empty dyad, y_{ij}^-) to a 1 (edge, y_{ij}^+), holding the rest of the network fixed: $y_{-ij} = y \setminus \{y_{ij}\}$. Strauss and Ikeda (1990) applied the pseudo-likelihood method of Besag (1977) to social networks which aims to approximate the full joint distribution in Equation 1 by the product of the full conditional probabilities of the network dyads:

$$f_{\text{PL}}(y | \theta) = \prod_{i \neq j \cup i < j} p(y_{ij} | y_{-ij}, \theta), \quad (2)$$

where the conditions $i \neq j$ and $i < j$ hold for directed and undirected networks, respectively. The pseudo-likelihood is equivalent to a logistic regression model (Wasserman and Pattison 1996) and ignores strong dependencies that might exist among network edges (e.g., transitivity effect) in the data: it can therefore lead to a biased estimation.

5. Parameter estimation

Bayesian analysis is a fully probabilistic treatment of uncertainty related to models and model parameters. A major advantage of the Bayesian approach is the flexibility with which prior information on the uncertainties related to the models and their parameters can be incorporated. In fact, the Bayesian approach permits the researcher to use both data and prior (e.g., expert-judgment) information in a consistent manner. For example, prior information about the data and/or from previous studies can easily be incorporated through an informative prior distribution (see for example Caimo, Pallotti, and Lomi 2017; Balest, Secco, Pisani, and Caimo 2019). This can be done by simply placing prior probability distributions on the possible values of the unknown parameters or models.

Let $p(\theta)$ be the prior distribution for the model parameters, θ . The posterior distribution of the parameters given the data can be obtained by using the Bayes' theorem:

$$\pi(\theta | y) = \frac{f(y | \theta) p(\theta)}{\pi(y)},$$

where $\pi(y) = \int_{\Theta} f(y | \theta) p(\theta) d\theta$ is the normalizing function for the posterior distribution, termed marginal likelihood or model evidence. In the ERGM context, both $z(\theta)$ and $\pi(y)$ are typically intractable and the posterior distribution is therefore computationally doubly-intractable.

5.1. The approximate exchange algorithm

In order to approximate the posterior distribution $\pi(\theta | y)$, the **Bergm** package uses the exchange algorithm described in Section 4.1 of [Caimo and Friel \(2011\)](#) to sample from the following distribution:

$$\pi(\theta', y', \theta | y) \propto f(y | \theta) p(\theta) \epsilon(\theta' | \theta) f(y' | \theta')$$

where $f(y' | \theta')$ is the likelihood on which the simulated data y' are defined and belongs to the same exponential family of densities as $f(y | \theta)$, $\epsilon(\theta' | \theta)$ is any arbitrary proposal distribution for the augmented variable θ' . At each MCMC iteration, the exchange algorithm consists of a Gibbs update of θ' followed by a Gibbs update of y' , which is drawn from the $f(\cdot | \theta')$ via an MCMC algorithm ([Hunter et al. 2008b](#)). Then an exchange or swap from the current state θ to the proposed new parameter θ' is performed.

Parallel adaptive direction sampler

In order to improve mixing a parallel adaptive direction sampler (ADS, [Gilks, Roberts, and George 1994](#); [Roberts and Gilks 1994](#)) is considered: at the i -th iteration of the algorithm we have a collection of H different chains interacting with one another. By construction, the state space consists of $\{\theta_1, \dots, \theta_H\}$ with target distribution $\pi(\theta_1 | y) \otimes \dots \otimes \pi(\theta_H | y)$. A parallel ADS move consists of generating a new parameter value θ'_h from the difference of two parameters θ_{h_1} and θ_{h_2} (randomly selected from other chains) multiplied by a scalar term γ (called ADS move factor) plus a random proposal ϵ (which is a multivariate Normal distribution in `bergm()`).

The format of the model specification is the same as the **ergm** package formula (to get a complete list of network statistics implemented in the **ergm** package use `?ergm-terms`). Let us consider the Lazega's law Firm network and a model including the following network statistics:

```
R> m1 <- lazega ~ edges + nodematch("Office") + nodematch("Practice") +
+   gwesp(0.5, fixed = TRUE)
```

In this case, our focus is on the density effect captured by the number of edges (`edges`), the homophily effect between lawyers working in the same office (`nodematch("Office")`) and in the same practice area (`nodematch("Practice")`), and the transitivity effect captured by the geometrically weighted edgewise shared partners statistic (GWESP) with fixed decay parameter equal to 0.5 (`gwesp(0.5, fixed = TRUE)`) ([Snijders et al. 2006](#)).

As mentioned in Section 5, we can specify prior distributions for the parameters in the model. The `bergm()` function allows users to specify the mean vector and variance/covariance matrix of a multivariate Normal distribution. For example we can create the `M.prior` object where we set the mean for the first parameter (corresponding to the `edge` statistic) to be equal to -4 which corresponds to assuming a priori that the average conditional odds of an edge between any two nodes i and j is $\exp(-4) \approx 0.018$. This reflects our prior assumption of overall sparsity of the network. The other prior mean values are set to be positive reflecting our prior assumption of positive homophily effect and transitivity as generally observed in this kind of social networks. The `S.prior` object is set to be a diagonal matrix with variances equal to 4.

```
R> M.prior <- c(-4, 0.5, 0.5, 1)
R> S.prior <- diag(4, 4)
```

By adopting the parallel ADS procedure we need to set the number of parallel chains by using the argument `nchains`. The number of chains must be greater than 3 and it is by default set to be twice the model dimension. For each chain, we can then set the number of burn-in iterations (`burn.in`) and the number of iterations after the burn-in (`main.iters`). The number of iterations used to simulate a network y' at each iteration is defined by the argument `aux.iters`. It is recommended to set `aux.iters` to be several times the number of dyads in the network so as to reduce the MCMC sampling error.

The arguments `prior.mean` and `prior.sigma` allow us to specify the multivariate Normal parameter prior distribution defined above.

The total number of iterations, i.e., the size of the posterior sample, is `nchains` \times `main.iters`. The proposal covariance structure of the proposal distribution $\epsilon(\cdot)$ is defined by the argument `V.proposal` which by default is set to be a diagonal matrix with every diagonal entry equal to 0.0025. In many cases, good mixing of the chain is ensured by a sensible tuning of the parallel ADS move factor `gamma` and therefore the argument `V.proposal` can be generally left at its default value. The parameter `gamma` can be easily tuned to achieve a suitable acceptance rate ($\sim 20\%$) by starting from its default value (0.5): the higher the value of `gamma` the lower the acceptance rate and vice versa. The range of values that `gamma` can take depends on the size of network and the kind of network statistics included in the model. The **Bergm** functions can only estimate ERGMs with dimensions greater than 1, meaning that at least 2 network statistics must be included in the model specification.

```
R> set.seed(1)
R> p.m1 <- bergm(m1, prior.mean = M.prior, prior.sigma = S.prior, burn.in = 500,
+   main.iters = 3000, aux.iters = 2500, nchains = 8, gamma = 0.6)
```

It is possible to summarize the posterior results of the MCMC estimation procedure by using the `summary()` function.

```
R> summary(p.m1)
```

```
Posterior Density Estimate for Model: y ~ edges + nodematch("Office") +
  nodematch("Practice") + gwesp(0.5, fixed = TRUE)
```

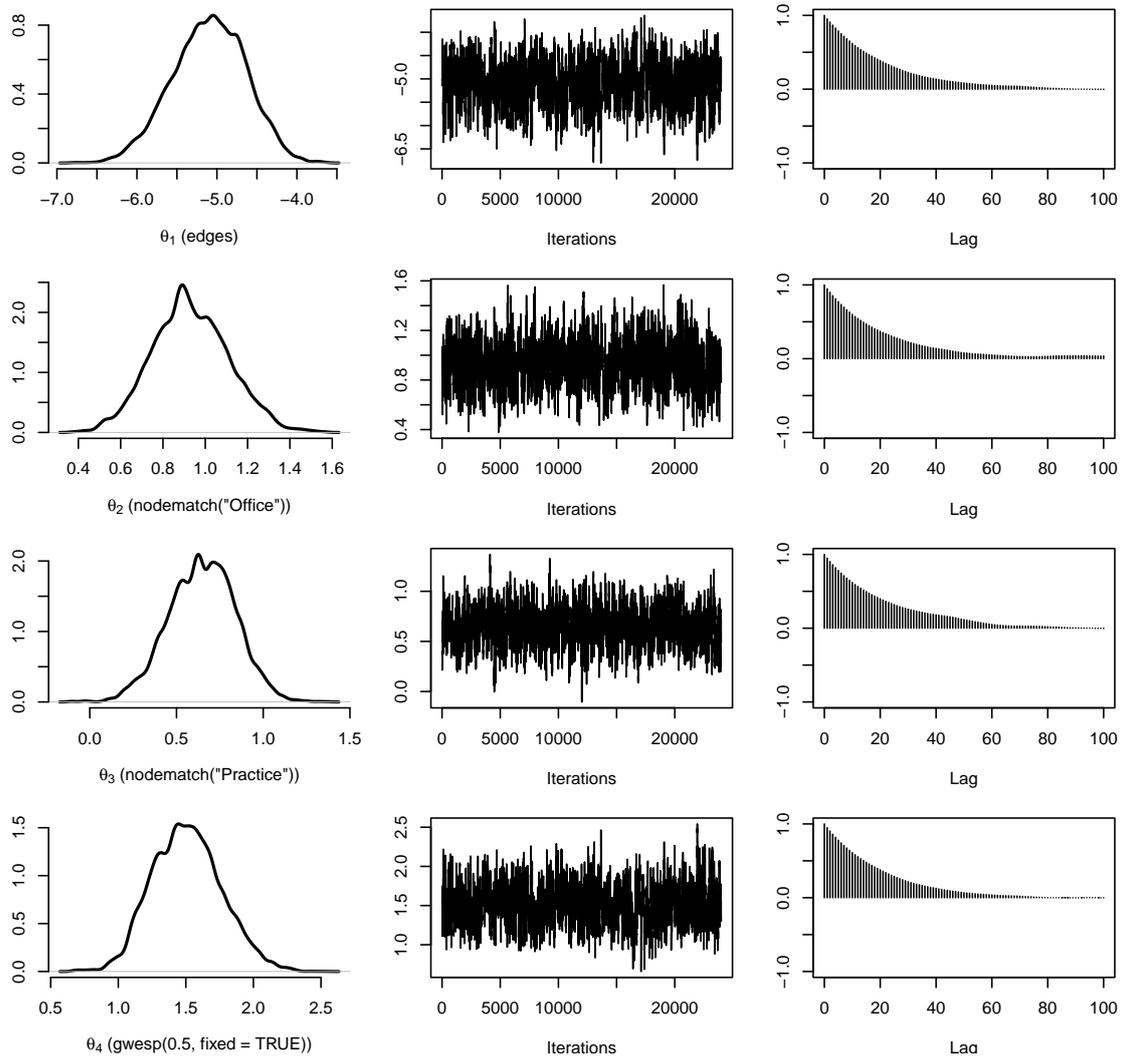


Figure 3: MCMC diagnostics for model p.m1.

	Mean	SD	Naive SE	Time-series SE
theta1 (edges)	-5.0984	0.4569	0.002949	0.019210
theta2 (nodematch.Office)	0.9310	0.1803	0.001164	0.007332
theta3 (nodematch.Practice)	0.6472	0.1915	0.001236	0.007827
theta4 (gwap.fixed.0.5)	1.5093	0.2544	0.001642	0.010416
	2.5%	25%	50%	75% 97.5%
theta1 (edges)	-6.0246	-5.4075	-5.0813	-4.7687 -4.244
theta2 (nodematch.Office)	0.5882	0.8077	0.9214	1.0521 1.295
theta3 (nodematch.Practice)	0.2585	0.5176	0.6484	0.7799 1.009
theta4 (gwap.fixed.0.5)	1.0570	1.3259	1.5035	1.6746 2.031

Acceptance rate: 0.19

The output above shows the results of the MCMC estimation: posterior means, standard deviations, medians, posterior quantile values and overall acceptance rate. In this example the 95% credible interval for the `edges` parameter θ_1 lies in the negative region whereas the 95% credible intervals for all the other parameters $\theta_2, \theta_3, \theta_4$ lie in positive regions. This means that the baseline edge probability is low and most of the edges of the network tend to connect nodes within the same office and in the same practice area and they tend to form triadic relations.

Figure 3 displays the MCMC diagnostic plots produced by the `plot()` function. The argument `lag` allows us to set the maximum lag for which autocorrelation is computed. The overall acceptance rate is 20% and the autocorrelation is negligible after lag 60; the trace plots indicate good mixing of the MCMC algorithm.

```
R> plot(p.m1, lag = 100)
```

5.2. Missing data augmentation

The estimation algorithm described above can lead to biases if y is not fully observed, because $s(y)$ cannot be properly calculated. A solution for this problem was proposed by Koskinen *et al.* (2010) and has been evaluated by Krause, Huisman, Steglich, and Snijders (2020). Let I be an indicator matrix of whether a tie variable is observed or missing, with $I_{ij} = 1$ if y_{ij} is observed and $I_{ij} = 0$ if y_{ij} is missing. Further we use the convention that u represents the observed part of the data ($I_{ij} = 1$) and v represents the unobserved part of the data ($I_{ij} = 0$). Thus the network can be reassembled as $y = (u, v)$. With the given network we can define an observation model for I , $f(I | y, \zeta)$, which is a probability model for what is observed and what is not, depending on the network y and some statistical parameter ζ . We present the algorithm limited to the setting where I is known and fixed and all covariates are known and fixed; extensions for multiplex networks exist (Krause and Caimo 2019). Further, the algorithm assumes that the missing data is ignorable, meaning the probability for data to be missing is independent from the missing values themselves and only dependent on the observed data. In other words, observations are “missing at random” (or “missing completely at random”) Rubin (1987). Thus we can ignore the missing data mechanism ζ in the estimation process. We now use data augmentation to estimate θ under missing data by alternating between draws from $\theta | u, v$ and $v | u, \theta$. We augment the observed data u by draws from the unobserved data v from the full conditional posterior $v^* \leftarrow v | u, \theta$, creating the augmented network $y^* = (u, v^*)$.

This process is implemented in `bergmM()` by adapting the approximate exchange algorithm presented above in the following way: an additional step to the MCMC procedure at each iteration is included if the proposed parameter vector θ' is accepted. In this additional step a new draw from $v^* \leftarrow v | u, I, \theta'$ is obtained and the augmented network y^* is updated. After which y^* is fed back to the algorithm as starting point for the next iteration as new baseline network. The augmented network is used to obtain the sufficient statistics $s(y^*)$, thus the third step of the algorithm does not optimize $s(y') - s(y)$, but $s(y') - s(y^*)$. A naive imputation is used for the first iteration of the algorithm to obtain an augmented starting network. In first step we substitute $s(y^*)$ with $s(u)$.

The algorithm can be summarized as shown in Algorithm 1.

The `bergmM()` function is similar to `bergm()` with two additional arguments, `nImp` and

Algorithm 1

Use a naive imputation to initialize $s(y^*)$
 Initialize θ
for $k = 1$ **to** K **do**
 Generate θ' using the ADS proposal procedure
 Simulate y' from $f(\cdot | \theta')$
 Update $\theta \rightarrow \theta'$ with the log of the probability:

$$\log \alpha = \min \left(0, [\theta - \theta']^\top [s(y') - s(y^*)] + \log \left[\frac{p(\theta')}{p(\theta)} \right] \right)$$

if θ' accepted **then**
 Simulate v^* from $f(\cdot | \theta', y^*)$ and generate a new $y^* = (u, v^*)$
end if
end for

`missingUpdate`. The argument `nImp` can be used to retain a specified number of imputed networks y^* from the estimation procedure. By default, imputed networks will not be returned. If more than two networks are to be retained during the estimation (`nImp > 2`), `bergmM()` will automatically space the sampling of y^* maximally over all iterations in `main.iters`. The second argument, `missingUpdate`, specifies how many tie swaps of the missing tie variables are simulated to obtain v^* from $f(\cdot | \theta', y^*)$ for each update of y^* . By default this is set to the number of missing edges, `missingUpdate = sum(is.na(y))`.

Below we illustrate the use of `bergmM()` on the Lazega's law firm network. Lazega's data is fully observed, thus we randomly set all outgoing ties of 4 nodes (11%) to missing. The CPU time for the `bergmM()` to complete the estimation is of the order of minutes.

```
R> set.seed(1)
R> missV <- sample(1:36, 4)
R> lazega[missV, ] <- lazega[, missV] <- NA
R> p.m1.M <- bergmM(m1, prior.mean = M.prior, prior.sigma = S.prior,
+   burn.in = 200, main.iters = 3000, aux.iters = 3000, nchains = 8,
+   gamma = 0.6, nImp = 10)
```

The object returned by `bergmM()` is a `bergm` object, thus the regular functions for assessing the estimation, obtaining MCMC diagnostics, and goodness of fit (see Section 6) apply.

```
R> summary(p.m1.M)
```

```
Posterior Density Estimate for Model: y ~ edges + nodematch("Office") +
  nodematch("Practice") + gwesp(0.5, fixed = TRUE)
```

	Mean	SD	Naive SE	Time-series SE
theta1 (edges)	-5.1599	0.4511	0.002912	0.019495
theta2 (nodematch.Office)	0.7903	0.1908	0.001232	0.008157
theta3 (nodematch.Practice)	0.5527	0.2014	0.001300	0.008550
theta4 (gwesp.fixed.0.5)	1.5903	0.2541	0.001640	0.011043

	2.5%	25%	50%	75%	97.5%
theta1 (edges)	-6.1095	-5.4350	-5.1462	-4.8603	-4.3220
theta2 (nodematch.Office)	0.4352	0.6562	0.7845	0.9159	1.1903
theta3 (nodematch.Practice)	0.1384	0.4230	0.5534	0.6886	0.9316
theta4 (gwesp.fixed.0.5)	1.1234	1.4104	1.5748	1.7434	2.1397

Acceptance rate: 0.18

In this example, the estimated parameter posterior density summaries obtained by the missing data estimation procedure are consistent with the ones obtained in Section 5.1.1 with no missing data. It is possible to get the imputed networks by typing:

```
R> p.m1.M$impNets
```

5.3. Fixing parameters during estimation

Sometimes it is necessary for theoretical or practical reasons to fix one or more of the model parameters during estimation. For instance, [Krivitsky, Handcock, and Morris \(2011\)](#) have shown that using a fixed offset term during ERGM estimation can reduce the differences in parameter estimates which can result from networks having different sizes. Similar to the **ergm** package ([Hunter *et al.* 2008b](#)), parameters that should be fixed during estimation need to be given as `offset()` in the model and a vector with values to which these parameters are to be fixed needs to be provided with the `offset.coef` argument. Note that one still needs to provide prior values for the fixed parameters. These values are, however, ignored during the estimation. Further, neither positive (`Inf`) nor negative infinity (`-Inf`) are valid inputs for `offset.coef`, instead one should use extreme numeric values (e.g., for the `mutual` parameter values of 100 or -100 can be considered as extreme). This feature is currently only implemented for the `bergm()` and `bergmM()` functions.

6. Goodness-of-fit diagnostics

[Hunter, Goodreau, and Handcock \(2008a\)](#) proposed systematic simulation-based goodness of fit (GOF) diagnostics for ERGMs, comparing several high-level statistics of observed networks with those of corresponding networks simulated from the estimated model. In the Bayesian framework, in order to evaluate the model goodness of fit in terms of posterior predictive assessment, the observed network is compared to a set of networks simulated from the estimated posterior distribution of the parameters of the model ([Caimo and Friel 2011](#)).

The `bgof()` function is used to carry out the Bayesian goodness-of-fit diagnostic procedure. The observed network is compared with a randomly simulated network sample (which size determined by the argument `sample.size`) drawn from the estimated posterior distribution using `aux.iters` iterations for the network simulation step.

In [Figures 4](#) the red lines represent the observed network GOF statistic values, the boxplots represent the GOF statistics of the simulated networks.

The `bgof()` function will produce the GOF diagnostic plots according to the type of network observed (directed or undirected). Depending on the type of network the user can specify

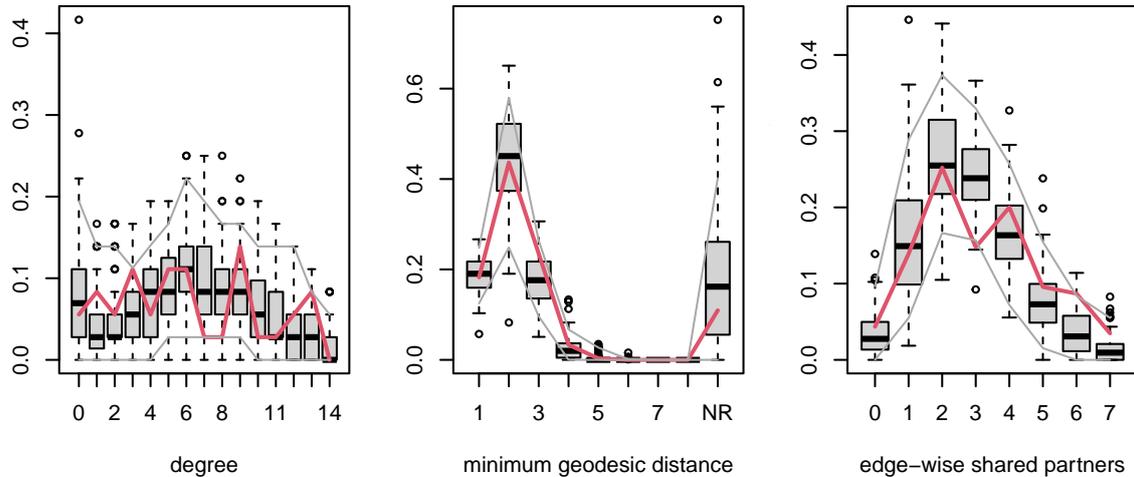


Figure 4: Bayesian goodness-of-fit diagnostics for the estimated parameter posterior distribution of model `p.m1`.

the maximum number of GOF distributions to be plotted. The set of statistics used for the comparison of directed graphs (such as the Lazega network) includes the degree distribution, the minimum geodesic distance distribution and the edgewise shared partner distribution. The arguments `n.deg`, `n.dist`, and `n.esp` indicate the number of boxplots to plot starting from the minimum value for each distribution, respectively.

```
R> set.seed(1)
R> bgof(p.m1, aux.iters = 5000, n.deg = 15, n.dist = 9, n.esp = 8)
```

Figure 4 shows that the networks simulated from the estimated posterior distributions are in reasonable agreement with the observed network as the red line is almost always falling inside the 95% interval represented by the light grey lines in the GOF plots.

7. Model selection

A model selection (or a model choice, or a model comparison) problem is faced when several competing statistical models are considered, any of which could serve as an explanation for our data, and we would like to select the best of them. In the ERGM context, this task translates into the choice of which subset of network statistics should be included in the model (Caimo and Friel 2013). Let us assume a countable model set $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \dots\}$. Each model indexed by m is defined by a likelihood function $f(y | \theta_m, \mathcal{M}_m) = f_m(y | \theta_m)$ and a prior on the model-specific parameter vector $\theta_m \in \Theta_m \subseteq \mathbb{R}^{d_m}$ denoted as $p(\theta_m | \mathcal{M}_m) = p_m(\theta_m)$, where d_m is the dimension of the parameter space Θ_m . The prior beliefs for each model are expressed through a prior distribution $p(\mathcal{M}_m)$, such that $\sum_{m=1}^{|\mathcal{M}|} p(\mathcal{M}_m) = 1$ where $|\mathcal{M}|$ is the cardinality of the model set.

Pairwise model comparisons can be performed with the posterior odds ratio between two

models \mathcal{M}_m and $\mathcal{M}_{m'}$:

$$\frac{\pi(\mathcal{M}_m | y)}{\pi(\mathcal{M}_{m'} | y)} = \frac{\pi(y | \mathcal{M}_m)}{\pi(y | \mathcal{M}_{m'})} \times \frac{p(\mathcal{M}_m)}{p(\mathcal{M}_{m'})}, \quad (3)$$

where $\frac{\pi(y | \mathcal{M}_m)}{\pi(y | \mathcal{M}_{m'})}$ is the Bayes factor;

$$\pi(\mathcal{M}_m | y) = \frac{\pi(y | \mathcal{M}_m)p(\mathcal{M}_m)}{\sum_{j=1}^{|\mathcal{M}|} \pi(y | \mathcal{M}_j)p(\mathcal{M}_j)}$$

is the posterior model probability for model \mathcal{M}_m ; and

$$\pi(y | \mathcal{M}_m) = \int_{\Theta_m} f_m(y | \theta_m)p_m(\theta_m) d\theta_m \quad (4)$$

is the model evidence or marginal likelihood under model \mathcal{M}_m .

The **Bergm** package assumes a multivariate Normal prior $\mathcal{N}_{d_m}(\mu_m, \Sigma_m)$ for θ_m , that leads to a marginal likelihood which is finite. Only for some elementary cases can the integration in Equation 4 be evaluated analytically. For moderately high-dimensional problems, Equation 4 is usually computationally intractable, and therefore sophisticated computational methods based on simulation are usually used to estimate it. There are many approaches based solely on within-model simulation that provide marginal likelihood estimates by utilizing the posterior samples of separate models. Recent reviews comparing popular methods based on MCMC sampling (all of which assume a tractable likelihood) can be found in [Friel and Wyse \(2012\)](#) and in [Ardia, Baştürk, Hoogerheide, and Van Dijk \(2012\)](#).

8. Pseudo-likelihood adjustment for large networks

Motivated by the inefficiency of the correction procedure of [Bouranis *et al.* \(2017\)](#) for model comparison (see `?bergmC` for more details), [Bouranis *et al.* \(2018\)](#) presented novel methodology for adjusting the pseudo-likelihood function directly as a means to obtain a reasonable and tractable approximation to the likelihood. These adjustments involve a correction of the mode, the curvature and the magnitude at the mode of the pseudo-likelihood in Equation 2 and are implemented in the `ergmAPL()` function. In this function, the model-specific fully adjusted pseudo-likelihood

$$\tilde{f}_m(y | \theta_m) = C_m \cdot f_{PL,m}(y | \hat{\theta}_{MPLE,m} + Q_m(\theta_m - \hat{\theta}_{MLE,m})), \quad (5)$$

depends on the maximum likelihood estimate, $\hat{\theta}_{MLE,m}$, the maximum pseudo-likelihood estimate, $\hat{\theta}_{MPLE,m}$, an upper triangular matrix Q_m of order d_m and the magnitude adjustment constant $C_m > 0$. Most crucially, Equation 5 renders the corresponding posterior distribution

$$\tilde{\pi}(\theta | y, \mathcal{M}_m) = \frac{\tilde{f}(y | \theta_m, \mathcal{M}_m)p(\theta_m | \mathcal{M}_m)}{\tilde{\pi}(y | \mathcal{M}_m)} = \frac{\tilde{f}(y | \theta_m, \mathcal{M}_m)p(\theta_m | \mathcal{M}_m)}{\int_{\Theta_m} \tilde{f}(y | \theta_m, \mathcal{M}_m)p(\theta_m | \mathcal{M}_m) d\theta_m} \quad (6)$$

amendable to standard evidence estimation methods from the Bayesian toolbox, allowing for Bayesian model selection of ERGMs.

In **Bergm**, the `evidence()` function estimates $\tilde{\pi}(y \mid \mathcal{M}_m)$ in Equation 6 using the Chib-Jeliazkov method (Chib and Jeliazkov 2001) or the power posterior method (Friel and Pettitt 2008; Friel, Hurn, and Wyse 2014).

Let us consider the Faux Dixon High School network and specify a higher dimensional model together with its parameter prior distribution. We assume that the model, labeled \mathcal{M}_1 , is almost identical to the model used to generate the simulated data (see `?faux.dixon.high`) and includes density (`edges`), mutuality (`mutual`) and transitivity (`gwesp`) effects, plus homophily effects for `race`, `sex` and `grade`; the number of nodes of in-degree 0 and 1 and the number of nodes of out-degree 0 and 1. We also assume that the normal parameter prior means are centered at 0 except for the `edges` parameter which is centered at -5, and that the variance/covariance matrix `S.prior1` is diagonal with entries equal to 5.

```
R> m1 <- dixon ~ edges + mutual + absdiff("grade") +
+   nodefactor("race") + nodefactor("grade") + nodefactor("sex") +
+   nodematch("race", diff = TRUE, levels = c("B", "O", "W")) +
+   nodematch("grade", diff = TRUE) + nodematch("sex", diff = FALSE) +
+   idegree(0:1) + odegree(0:1) + gwesp(0.1, fixed = TRUE)
R> M.prior1 <- c(-5, rep(0, 26))
R> S.prior1 <- diag(5, 27)
```

The `evidence()` function is used to carry out MCMC sampling from the posterior distribution of Equation 6. The total number of iterations per chain is `burn.in + main.iters`, with the first `burn.in` MCMC draws removed from the posterior sample. The covariance structure of the multivariate Normal proposal distribution for the MCMC run can be modified by the argument `V.proposal`, which can be easily tuned to achieve a suitable acceptance rate by starting from its default value (1.5). The approximate contrastive divergence (CD) procedure (specified by the `estimate` argument) is faster but less accurate than the approximate maximum likelihood procedure for $\hat{\theta}_{MLE,m}$, as the function `ergm()` that is called from the `ergmAPL()` function may require several minutes to estimate the $\hat{\theta}_{MLE,m}$, due to the large size of the network and the higher dimensionality of the parameter space. The importance sampling algorithm for estimating the magnitude adjustment constant $C_m > 0$ was carried out using `ladder = 200` path points, `aux.iters = 2500` auxiliary iterations used for drawing the first network from the ERGM likelihood at each iteration, `n.aux.draws = 50` auxiliary networks drawn from the ERGM likelihood and `aux.thin = 50` auxiliary iterations between each of the `n.aux.draws` network draws after the first network is drawn.

```
R> cj1 <- evidence(evidence.method = "CJ", formula = m1,
+   prior.mean = M.prior1, prior.sigma = S.prior1, aux.iters = 2500,
+   n.aux.draws = 50, aux.thin = 50, ladder = 200, V.proposal = 0.5,
+   burn.in = 5000, main.iters = 30000, num.samples = 25000,
+   estimate = "CD", seed = 1)
```

The posterior summaries can be obtained by using the `summary()` function.

```
R> summary(cj1)
```

```
Posterior Density Estimate for Model: y ~ edges + mutual +
  absdiff("grade") + nodefactor("race") +
```

```

nodefactor("grade") + nodefactor("sex") +
nodematch("race", diff = TRUE, levels = c("B", "O", "W")) +
nodematch("grade", diff = TRUE) + nodematch("sex", diff = FALSE) +
idegree(0:1) + odegree(0:1) +
gwesp(0.1, fixed = TRUE)

```

	Mean	SD	Naive SE	Time-series SE
theta1 (edges)	-4.813061	1.5823	0.010007	0.09669
theta2 (mutual)	1.068195	1.4147	0.008947	0.07934
theta3 (absdiff.grade)	-0.508263	0.3875	0.002451	0.02156
theta4 (nodefactor.race.H)	0.459097	1.2733	0.008053	0.06932
theta5 (nodefactor.race.O)	0.799284	1.0462	0.006617	0.06150
theta6 (nodefactor.race.W)	-0.263883	0.9075	0.005739	0.05285
theta7 (nodefactor.grade.8)	-0.089955	0.8364	0.005290	0.04892
theta8 (nodefactor.grade.9)	0.416923	0.7980	0.005047	0.04691
theta9 (nodefactor.grade.10)	-0.062195	0.9215	0.005828	0.05516
theta10 (nodefactor.grade.11)	0.273094	0.9639	0.006096	0.05493
theta11 (nodefactor.grade.12)	0.443116	0.8997	0.005690	0.05316
theta12 (nodefactor.sex.2)	-0.250685	0.3456	0.002186	0.01953
theta13 (nodematch.race.B)	0.829879	1.3518	0.008549	0.08131
theta14 (nodematch.race.O)	0.161942	2.1373	0.013518	0.11985
theta15 (nodematch.race.W)	1.033112	1.1838	0.007487	0.07116
theta16 (nodematch.grade.7)	0.308318	1.7322	0.010955	0.10153
theta17 (nodematch.grade.8)	0.381694	1.4582	0.009222	0.08130
theta18 (nodematch.grade.9)	-0.276393	1.3611	0.008608	0.07937
theta19 (nodematch.grade.10)	0.894538	1.4561	0.009209	0.08502
theta20 (nodematch.grade.11)	0.005533	1.7994	0.011381	0.10349
theta21 (nodematch.grade.12)	0.140412	1.5670	0.009911	0.09003
theta22 (nodematch.sex)	-0.106814	0.6738	0.004262	0.03779
theta23 (idegree0)	0.222886	1.4779	0.009347	0.08291
theta24 (idegree1)	0.208954	1.0037	0.006348	0.05965
theta25 (odegree0)	1.084426	1.7162	0.010854	0.10026
theta26 (odegree1)	-0.122699	1.1425	0.007226	0.06557
theta27 (gwesp.fixed.0.1)	1.391942	0.5287	0.003344	0.03133

	2.5%	25%	50%	75%	97.5%
theta1 (edges)	-7.8453	-5.90827	-4.81718	-3.71903	-1.7342
theta2 (mutual)	-1.7248	0.15401	1.04661	1.98236	3.9298
theta3 (absdiff.grade)	-1.2518	-0.77787	-0.51708	-0.25225	0.2515
theta4 (nodefactor.race.H)	-1.9633	-0.41868	0.47257	1.28103	3.0344
theta5 (nodefactor.race.O)	-1.2563	0.09420	0.79495	1.54599	2.7617
theta6 (nodefactor.race.W)	-1.9805	-0.88429	-0.28089	0.29370	1.6247
theta7 (nodefactor.grade.8)	-1.6980	-0.63370	-0.10016	0.46475	1.5430
theta8 (nodefactor.grade.9)	-1.1419	-0.13165	0.42863	0.96269	1.9412
theta9 (nodefactor.grade.10)	-1.9713	-0.65783	-0.05612	0.55868	1.6874
theta10 (nodefactor.grade.11)	-1.5681	-0.41096	0.28435	0.92284	2.2124
theta11 (nodefactor.grade.12)	-1.3811	-0.17253	0.45345	1.04748	2.1648

theta12 (nodefactor.sex.2)	-0.8996	-0.48241	-0.26409	-0.01418	0.4563
theta13 (nodematch.race.B)	-1.7780	-0.09994	0.80456	1.74570	3.5177
theta14 (nodematch.race.O)	-4.1354	-1.27457	0.17541	1.57718	4.3670
theta15 (nodematch.race.W)	-1.2894	0.24117	1.00636	1.83519	3.3785
theta16 (nodematch.grade.7)	-3.1801	-0.88409	0.30731	1.47915	3.7342
theta17 (nodematch.grade.8)	-2.4807	-0.55112	0.34486	1.34164	3.1812
theta18 (nodematch.grade.9)	-2.9888	-1.20582	-0.24977	0.65012	2.3832
theta19 (nodematch.grade.10)	-1.8870	-0.11321	0.85152	1.83439	3.8801
theta20 (nodematch.grade.11)	-3.6269	-1.20096	0.02556	1.23647	3.5722
theta21 (nodematch.grade.12)	-3.0377	-0.88408	0.16743	1.22917	3.0526
theta22 (nodematch.sex)	-1.3460	-0.58812	-0.12329	0.33952	1.2882
theta23 (idegree0)	-2.6020	-0.81089	0.16693	1.26942	3.1648
theta24 (idegree1)	-1.7344	-0.48650	0.21102	0.90154	2.1715
theta25 (odegree0)	-2.2439	-0.11685	1.08229	2.25297	4.4425
theta26 (odegree1)	-2.3276	-0.90311	-0.12563	0.64406	2.1012
theta27 (gwesp.fixed.0.1)	0.3722	1.02480	1.39944	1.76710	2.3987

Acceptance rate: 0.21

Figure 5 displays the MCMC diagnostic plots produced by the `plot()` function for the first four model parameters. The overall acceptance rate is about 20% and the autocorrelation is negligible after lag 100. The posterior parameter estimates obtained by the `evidence()` function can be assessed using the `bgof()` function.

```
R> plot(cj1)
```

The set of GOF statistic distributions used for the comparison of directed graphs (such as the Dixon network) includes the in-degree distribution, the out-degree distribution, the minimum geodesic distance distribution and the edgewise shared partner distribution. The arguments `n.ideg`, `n.odeg`, `n.dist`, and `n.esp` indicate the number of boxplots to plot for each distribution, respectively.

```
R> set.seed(1)
```

```
R> bgof(cj1, sample.size = 100, aux.iters = 5000, n.ideg = 20, n.odeg = 20,
+       n.dist = 10, n.esp = 7)
```

Figure 6 shows that the `cj1` model performs well in reproducing the observed global network properties that are not included in the model.

There are three main factors influencing the computational intensity of the estimation methods: the observed network size, the ERGM model size, and the computational complexity of the network statistics to include in the model. The adjusted pseudo-likelihood and pseudo-posterior procedures allow for fast approximate inference of the ERGM parameter posterior and model evidence for networks on several thousands nodes.

The `evidence()` function includes an additional feature that allows for estimation of the log-model evidence using the fully adjusted pseudo-likelihood; it is a wrapper for the functions `evidenceCJ()` and `evidencePP()`, which implement Chib and Jeliazkov's method and the power posteriors method, respectively. For this example, Chib and Jeliazkov's method was

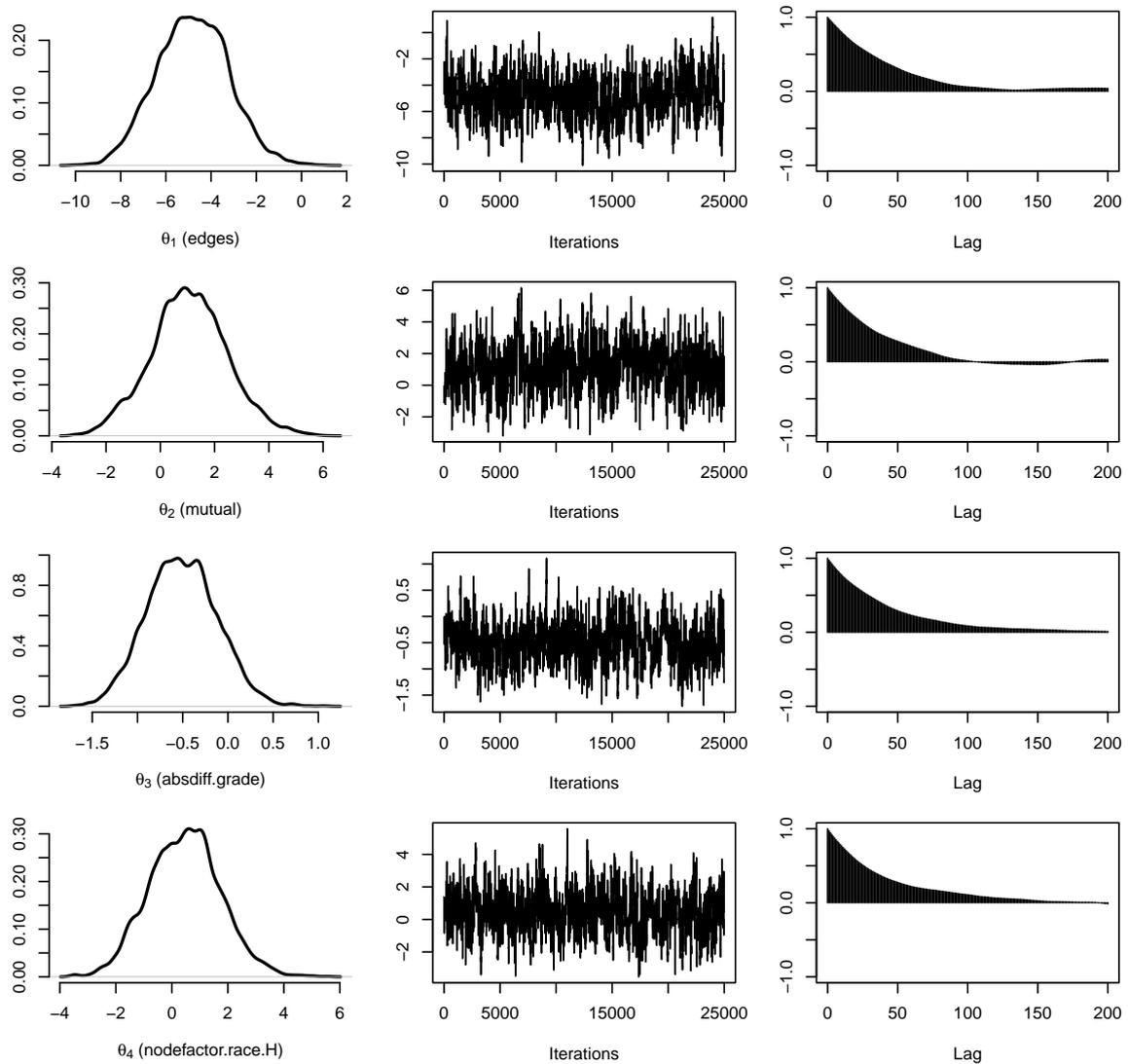


Figure 5: MCMC diagnostics for model `cj1` (first four parameters only).

selected with the option `evidence.method = "CJ"` and `num.samples = 25000` MCMC draws were kept for estimating the log-evidence.

We proceed with a model selection example by defining model \mathcal{M}_2 , which assumes that the terms for mutuality and transitivity are removed from \mathcal{M}_1 .

```
R> m2 <- dixon ~ edges + absdiff("grade") +
+   nodefactor("race") + nodefactor("grade") + nodefactor("sex") +
+   nodematch("race", diff = TRUE, levels = c("B", "O", "W")) +
+   nodematch("grade", diff = TRUE) + nodematch("sex", diff = FALSE) +
+   idegree(0:1) + odegree(0:1)
R> M.prior2 <- c(-5, rep(0, 24))
R> S.prior2 <- diag(5, 25)
```

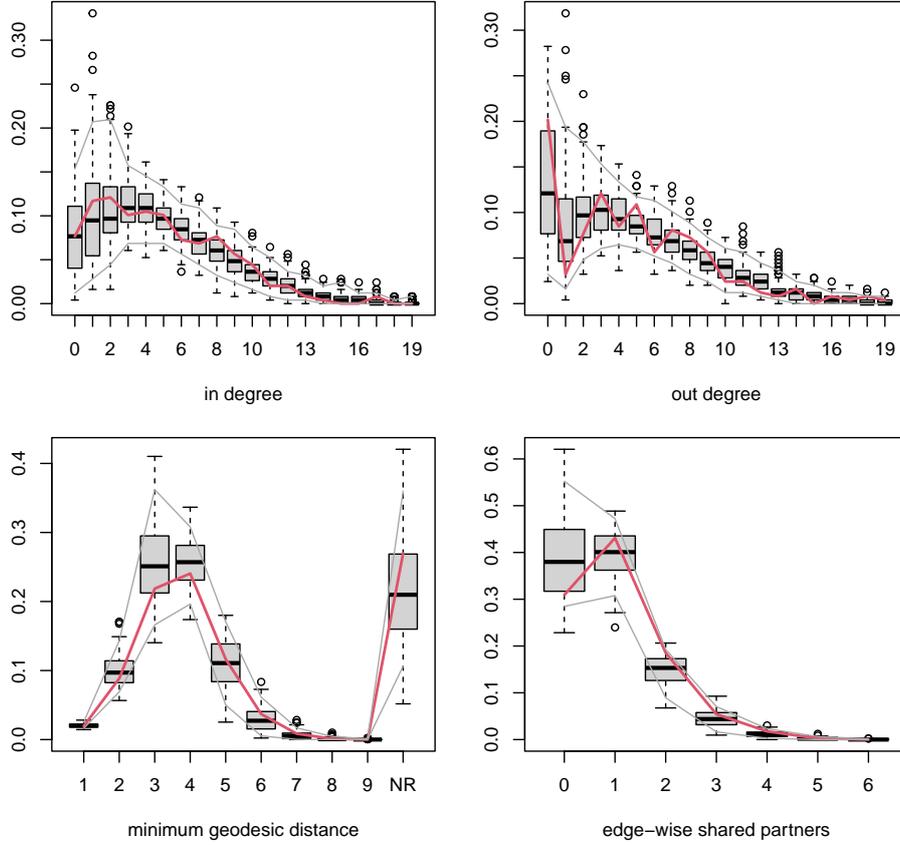


Figure 6: Bayesian goodness-of-fit diagnostics for the estimated parameter posterior distribution of model *cj1*.

Model	Log evidence estimate	CPU (mins)	BF ₁₂
\mathcal{M}_1	-38079.99	2.92	5.10×10^{55}
\mathcal{M}_2	-38208.26	1.61	

Table 1: Faux Dixon High School: Log evidence estimates, CPU time in minutes and resulting Bayes factor (BF) estimate for the models under consideration.

```
R> cj2 <- evidence(evidence.method = "CJ", formula = m2,
+   prior.mean = M.prior2, prior.sigma = S.prior2, aux.iters = 2500,
+   n.aux.draws = 50, aux.thin = 50, ladder = 200, V.proposal = 0.5,
+   burn.in = 5000, main.iters = 30000, num.samples = 25000,
+   estimate = "CD", seed = 1)
```

When two models are equally probable a priori, so that $p(\mathcal{M}_1) = p(\mathcal{M}_2)$, the Bayes factor is equal to the posterior odds ratio of \mathcal{M}_1 and \mathcal{M}_2 in Equation 3. In this example the estimated Bayes factor provides strong evidence in favor of model \mathcal{M}_1 , as expected (Table 1). This reveals that the transitivity and mutuality effect are important connectivity features of the observed network but also the homophily effect of **race**, **sex** and **grade** can help explain the complexity of the observed network data. Under this setting and the increased model

complexity, the CPU time for each implementation of Chib and Jeliazkov’s method is of the order of minutes (Table 1). The same estimation would require a few hours using the `bergm()` function.

9. Discussion

The software package **Bergm** aims to help researchers and practitioners in two ways. Firstly, it provides a simple, efficient and complete range of tools for conducting Bayesian inference for exponential random graph models. Secondly, **Bergm** makes available a platform that can be easily customized, extended, and adapted to address different requirements. The software package is under continual maintenance and periodic significant upgrading. Future developments will include uncertainty quantification of the Monte Carlo estimates of the evidence and extensions to weighted networks (Caimo and Gollini 2020).

The aim of this tutorial is to serve as a useful introduction to the main capabilities of the package as well as some of the algorithms and methods behind it.

References

- Alquier P, Friel N, Everitt R, Boland A (2016). “Noisy Monte Carlo: Convergence of Markov Chains with Approximate Transition Kernels.” *Statistics and Computing*, **26**(1-2), 29–47. doi:10.1007/s11222-014-9521-x.
- Ardia D, Baştürk N, Hoogerheide L, Van Dijk HK (2012). “A Comparative Study of Monte Carlo Methods for Efficient Evaluation of Marginal Likelihood.” *Computational Statistics & Data Analysis*, **56**, 3398–3414. doi:10.1016/j.csda.2010.09.001.
- Balest J, Secco L, Pisani E, Caimo A (2019). “Sustainable Energy Governance in South Tyrol (Italy): A Probabilistic Bipartite Network Model.” *Journal of Cleaner Production*, **221**, 854–862. doi:10.1016/j.jclepro.2019.02.191.
- Bates D, Mächler M (2022). **Matrix: Sparse and Dense Matrix Classes and Methods**. R package version 1.4.1, URL <https://CRAN.R-project.org/package=Matrix>.
- Besag J (1977). “Efficiency of Pseudolikelihood Estimation for Simple Gaussian Fields.” *Biometrika*, **64**, 616–618. doi:10.1093/biomet/64.3.616.
- Bouranis L, Friel N, Maire F (2017). “Efficient Bayesian Inference for Exponential Random Graph Models by Correcting the Pseudo-Posterior Distribution.” *Social Networks*, **50**, 98–108. doi:10.1016/j.socnet.2017.03.013.
- Bouranis L, Friel N, Maire F (2018). “Bayesian Model Selection for Exponential Random Graph Models via Adjusted Pseudolikelihoods.” *Journal of Computational and Graphical Statistics*, **27**(3), 516–528. doi:10.1080/10618600.2018.1448832.
- Box-Steffensmeier JM, Christenson DP, Morgan JW (2018). “Modeling Unobserved Heterogeneity in Social Networks with the Frailty Exponential Random Graph Model.” *Political Analysis*, **26**(1), 3–19. doi:10.1017/pan.2017.23.

- Butts CT (2008). “**network**: A Package for Managing Relational Data in R.” *Journal of Statistical Software*, **24**(2). doi:10.18637/jss.v024.i02.
- Caimo A, Bouranis L, Krause R (2022). **Bergm**: Bayesian Exponential Random Graph Models. R package version 5.0.5, URL <https://CRAN.R-project.org/package=Bergm>.
- Caimo A, Friel N (2011). “Bayesian Inference for Exponential Random Graph Models.” *Social Networks*, **33**(1), 41–55. doi:10.1016/j.socnet.2010.09.004.
- Caimo A, Friel N (2013). “Bayesian Model Selection for Exponential Random Graph Models.” *Social Networks*, **35**(1), 11–24. doi:10.1016/j.socnet.2012.10.003.
- Caimo A, Friel N (2014). “**Bergm**: Bayesian Exponential Random Graphs in R.” *Journal of Statistical Software*, **61**(2), 1–25. doi:10.18637/jss.v061.i02.
- Caimo A, Gollini I (2020). “A Multilayer Exponential Random Graph Modelling Approach for Weighted Networks.” *Computational Statistics & Data Analysis*, **142**, 106825. doi:10.1016/j.csda.2019.106825.
- Caimo A, Lomi A (2014). “Knowledge Sharing in Organizations: A Bayesian Analysis of the Role of Reciprocity and Formal Structure.” *Journal of Management*, **41**, 665–691. doi:10.1177/0149206314552192.
- Caimo A, Mira A (2015). “Efficient Computational Strategies for Doubly Intractable Problems with Applications to Bayesian Social Networks.” *Statistics and Computing*, **25**, 113–125. doi:10.1007/s11222-014-9516-7.
- Caimo A, Pallotti F, Lomi A (2017). “Bayesian Exponential Random Graph Modelling of Interhospital Patient Referral Networks.” *Statistics in Medicine*, **36**(18), 2902–2920. doi:10.1002/sim.7301.
- Chib S, Jeliazkov I (2001). “Marginal Likelihood from the Metropolis-Hastings Output.” *Journal of the American Statistical Association*, **96**, 270–281. doi:10.1198/016214501750332848.
- Frank O, Strauss D (1986). “Markov Graphs.” *Journal of the American Statistical Association*, **81**, 832–842. doi:10.1080/01621459.1986.10478342.
- Friel N, Hurn M, Wyse J (2014). “Improving Power Posterior Estimation of Statistical Evidence.” *Statistics and Computing*, **24**, 709–723. doi:10.1007/s11222-013-9397-1.
- Friel N, Pettitt AN (2008). “Marginal Likelihood Estimation via Power Posteriors.” *Journal of the Royal Statistical Society B*, **70**(3), 589–607. doi:10.1111/j.1467-9868.2007.00650.x.
- Friel N, Wyse J (2012). “Estimating the Evidence – A Review.” *Statistica Neerlandica*, **66**(3), 288–308. doi:10.1111/j.1467-9574.2011.00515.x.
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2021). **mvtnorm**: Multivariate Normal and *t* Distributions. R package version 1.1-3, URL <https://CRAN.R-project.org/package=mvtnorm>.

- Gilks WR, Roberts GO, George EI (1994). “Adaptive Direction Sampling.” *Statistician*, **43**(1), 179–189. doi:10.2307/2348942.
- Handcock M, Hunter D, Butts C, Goodreau S, Morris M (2008). “**statnet**: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data.” *Journal of Statistical Software*, **24**(1), 1–11. doi:10.18637/jss.v024.i01.
- Henning C, Aßmann C, Hedtrich J, Ehrenfels J, Krampe E (2019). “What Drives Participatory Policy Processes: Grassroot Activities, Scientific Knowledge or Donor Money? A Comparative Policy Network Approach.” *Social Networks*, **58**, 78–104. doi:10.1016/j.socnet.2019.03.001.
- Holland PW, Leinhardt S (1981). “An Exponential Family of Probability Distributions for Directed Graphs.” *Journal of the American Statistical Association*, **76**, 33–65. doi:10.1080/01621459.1981.10477598.
- Hunter D, Handcock M (2006). “Inference in Curved Exponential Family Models for Networks.” *Journal of Computational and Graphical Statistics*, **15**(3), 565–583. doi:10.1198/106186006x133069.
- Hunter DR, Goodreau SM, Handcock MS (2008a). “Goodness of Fit of Social Network Models.” *Journal of the American Statistical Association*, **103**(481), 248–258. doi:10.1198/016214507000000446.
- Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M (2008b). “**ergm**: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks.” *Journal of Statistical Software*, **24**(3), 1–29. doi:10.18637/jss.v024.i03.
- Koskinen JH, Robins GL, Pattison PE (2010). “Analysing Exponential Random Graph (p-Star) Models with Missing Data Using Bayesian Data Augmentation.” *Statistical Methodology*, **7**(3), 366–384. doi:10.1016/j.stamet.2009.09.007.
- Krause RW, Caimo A (2019). “Missing Data Augmentation for Bayesian Exponential Random Multi-Graph Models.” In *International Workshop on Complex Networks*, pp. 63–72. Springer.
- Krause RW, Huisman M, Steglich CEG, Snijders TAB (2020). “Missing Data in Cross-Sectional Networks – An Extensive Comparison of Missing Data Treatment Methods.” *Social Networks*, **62**, 99–112. doi:10.1016/j.socnet.2020.02.004.
- Krivitsky PN, Handcock MS (2014). “A Separable Model for Dynamic Networks.” *Journal of the Royal Statistical Society B*, **76**(1), 29–46. doi:10.1111/rssb.12014.
- Krivitsky PN, Handcock MS, Morris M (2011). “Adjusting for Network Size and Composition Effects in Exponential-Family Random Graph Models.” *Statistical Methodology*, **8**(4), 319–339. doi:10.1016/j.stamet.2011.01.005.
- Lazega E (2001). *The Collegial Phenomenon: The Social Mechanisms of Cooperation among Peers in a Corporate Law Partnership*. 1st edition. Oxford University Press.

- Leifeld P, Cranmer SJ, Desmarais BA (2018). “Temporal Exponential Random Graph Models with **btergm**: Estimation and Bootstrap Confidence Intervals.” *Journal of Statistical Software*, **83**(6), 1–36. doi:10.18637/jss.v083.i06.
- Martin AD, Quinn KM, Park JH (2011). “**MCMCpack**: Markov Chain Monte Carlo in R.” *Journal of Statistical Software*, **42**(9), 22. doi:10.18637/jss.v042.i09.
- Novomestky F (2012). **matrixcalc**: *Collection of Functions for Matrix Calculations*. R package version 1.0-5, URL <https://CRAN.R-project.org/package=matrixcalc>.
- Plummer M, Best N, Cowles K, Vines K (2006). “**coda**: Convergence Diagnosis and Output Analysis for MCMC.” *R News*, **6**(1), 7–11. URL <https://CRAN.R-project.org/doc/Rnews/>.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Resnick M, Bearman P, Blum R, Bauman K, Harris K, Jones J, Tabor J, Beuhring T, Sieving R, Shew M, Ireland M, Bearinger L, Udry J (1997). “Protecting Adolescents from Harm. Findings from the National Longitudinal Study on Adolescent Health.” *Journal of the American Medical Association*, **278**(10), 823–32. doi:10.1001/jama.1997.03550100049038.
- Roberts GO, Gilks WR (1994). “Convergence of Adaptive Direction Sampling.” *Journal of Multivariate Analysis*, **49**(2), 287–298. doi:10.1006/jmva.1994.1028.
- Robins G, Pattison P, Kalish Y, Lusher D (2007). “An Introduction to Exponential Random Graph Models for Social Networks.” *Social Networks*, **29**(2), 169–348. doi:10.1016/j.socnet.2006.08.002.
- Rubin DB (1987). *Multiple Imputation for Nonresponse in Surveys*. 1st edition. John Wiley & Sons. doi:10.1002/9780470316696.
- Sinke MRT, Dijkhuizen RM, Caimo A, Stam CJ, Otte WM (2016). “Bayesian Exponential Random Graph Modeling of Whole-Brain Structural Networks across Lifespan.” *Neuroimage*, **135**, 79–91. doi:10.1016/j.neuroimage.2016.04.066.
- Snijders T, Pattison P, Robins G, Handcock M (2006). “New Specifications for Exponential Random Graph Models.” *Sociological Methodology*, **36**, 99–153. doi:10.1111/j.1467-9531.2006.00176.x.
- Stewart J, Schweinberger M (2018). **mlergm**: *Multilevel Exponential-Family Random Graph Models*. R package version 0.1, URL <https://CRAN.R-project.org/package=mlergm>.
- Strauss D, Ikeda M (1990). “Pseudolikelihood Estimation for Social Networks.” *Journal of the American Statistical Association*, **85**, 204–212. doi:10.1080/01621459.1990.10475327.
- Tasselli S, Caimo A (2019). “Does It Take Three to Dance the Tango? Organizational Design, Triadic Structures and Boundary Spanning across Subunits.” *Social Networks*, **59**, 10–22. doi:10.1016/j.socnet.2019.04.002.
- Theussl S, Hornik K (2019). **Rglpk**: *R/GNU Linear Programming Kit Interface*. R package version 0.6-4, URL <https://CRAN.R-project.org/package=Rglpk>.

Wang P, Robins G, Pattison P (2009). *PNet: A Program for the Simulation and Estimation of Exponential Random Graph Models*. URL <http://www.melnet.org.au/pnet>.

Wasserman S, Pattison P (1996). “Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^* .” *Psychometrika*, **61**, 401–425. doi:10.1007/bf02294547.

Affiliation:

Alberto Caimo
School of Mathematics and Statistics
Technological University Dublin
Grangegorman, Dublin
D07 ADY7 Ireland
E-mail: alberto.caimo@tudublin.ie
URL: <https://acaimo.github.io/>