Deterministic Bayesian inference for the p^* model

Haakon Austad
Norwegian University of Science and Technology

Nial Friel
University College Dublin

Abstract

The p* model is widely used in social network analysis. The likelihood of a network under this model is impossible to calculate for all but trivially small networks. Various approximation have been presented in the literature, and the pseudolikelihood approximation is the most popular. The aim of this paper is to introduce two likelihood approximations which have the pseudolikelihood estimator as a special case. We show, for the examples that we have considered, that both approximations result in improved estimation of model parameters with respect to the standard methodological approaches. We provide a deterministic approach and also illustrate how Bayesian model choice can be carried out in this setting.

1 Introduction

Many probability models have been developed in order to summarise the general structure of networks. For example, the Bernoulli random graph model (Erdös and Rényi, 1959) assumes that edges are considered independent of each other; the p_1 model (Holland and Leinhardt, 1981) assumes independent edge variables. The Markov random graph model (Frank and Strauss, 1986) assumes that each pair of edges is conditionally dependent given the rest of the graph. The family of p^* or exponential random graph models (Wasserman and Pattison (1996), see also Robins et al. (2007) for a recent review) is a generalisation of the latter model and is thought to be a flexible way to model the complex dependence structure of network graphs. The p^* model is arguably the most widely used model in social network analysis.

Despite this popularity, the main drawback to the p^* model is that the likelihood is generally unavailable, since it involves a summation over $2^{m(m-1)/2}$ terms for a network with m nodes. Clearly the size of this summation grows super-exponentially with m. For this reason various approximations have been presented in the literature. The most widely used approximation is the pseudolikelihood estimator (Strauss and Ikeda, 1990) which dates back to (Besag, 1974). It is well understood that this approximation can give poor performance, for example in the context of the autologistic distribution (Friel et al., 2009). However no formal assessment of the performance the pseudolikelihood estimator in the context of the p^* model has yet been established, and a partial aim of this paper is to address this problem. The main contribution of this paper is to introduce and investigate the use of two likelihood approximations which have the pseudolikelihood estimator as a special case.

We consider a deterministic simulation-free inference approach, avoiding the need for Markov chain Monte Carlo methods, along the lines of Rue et al. (2009). Essentially, the dimension of the parameter space is usually quite small, often with 5 or less parameters. Therefore evaluating the unnormalised posterior distribution on a fine grid is possible. An advantage of this approach is that estimates of posterior model probabilities are then available for all models nested within a model of maximal dimension using a grid evaluated for the model of maximal dimension.

There are other approaches which one could take to carry out inference for the p^* model, using simulation methods, for example. A popular choice in this setting is the Monte Carlo MLE approach of Geyer and Thompson (1992). This involves an importance sampling estimator of a ratio of normalising constants for the different parameter values of the p^* model. This method turns out to be quite difficult to implement – it involves drawing graphs from the likelihood with pre-specified initial parameters. However the choice of initial parameters is crucial, since a poorly chosen initial parameters, lying in the degenerate region of the parameter, for example, may result in simulated

graphs which are empty or full. This in turn impacts negatively on the parameter estimation.

The paper is organised as follows. In section 2 we introduce the exponential random graph model. The new likelihood approximations are outlined in Section 3. While Section 5 demonstrates their performance in a number of examples, where we consider not only posterior estimation but also Bayesian model choice. Finally Section 6 offers some conclusions and discusses possible improvements to the methodology.

2   The exponential random graph model

Consider a random adjacency matrix $\boldsymbol{Y}$ representing a graph on m nodes. It can be defined by the set $\{Y_{ij} : i = 1, \ldots, m; j = 1, \ldots, m\}$ where the dyad $Y_{ij} = 1$ if the pair (i, j) is connected, and $Y_{ij} = 0$ otherwise. The diagonal entries of $\boldsymbol{y}$ take the value 0. The edges in the graphs could be either directed on undirected. In this study we have chosen to only look at undirected graphs, but similar techniques could be established for directed graphs as well. Let $\mathcal{Y}$ denote the set of all possible graphs on n nodes and let $\boldsymbol{y}$ be a realisation of $\boldsymbol{Y}$. The p^* model writes the probability distribution of $\boldsymbol{Y}$ as

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

where $s(\boldsymbol{y})$ is a known vector of sufficient statistics, for example,

$$
\begin{aligned}
&s_1(\boldsymbol{y}) = \sum_{i<j} y_{ij}, && \text{number of edges,} \\
&s_2(\boldsymbol{y}) = \sum_{i<j<k} y_{ik}y_{jk}, && \text{number of two-stars,} \\
&s_3(\boldsymbol{y}) = \sum_{i<j<k<l} y_{il}y_{jl}y_{kl}, && \text{number of three-stars,} \\
&s_4(\boldsymbol{y}) = \sum_{i<j<k} y_{jk}y_{ik}y_{ij} && \text{number of triangles.}
\end{aligned}
$$

Finally β are model parameters corresponding the collection of sufficient statistics. Formally the p^* model is a Markov random field, where two edges are neighbours of one another if they share a common node. A graph with m nodes contains $n = m(m-1)/2$ edges, each of which can take values 0 or 1. Thus $\mathcal{Y}$ contains 2^n possible undirected graphs and the normalising constant $z(\beta) = \sum_{\boldsymbol{y} \in \mathcal{Y}} \exp\{\beta^t s(\boldsymbol{y})\}$ is consequently extremely difficult to evaluate for all but trivially small graphs.

2.1   Model degeneracy

Model degeneracy is an important issue concerning p^* models and was largely treated in Handcock (2003) and more recently in Rinaldo et al. (2009). The term degeneracy refers to the fact that for a network with a given number of nodes, there are so-called degenerate regions of the parameter space from which simulated networks will be either empty or full (complete). In fact, the non-degenerate region is typically a very thin region in the parameter space. Model degeneracy presents a considerable challenge for parameter estimation. Consider the mean parameterisation for the p^* model defined by $\mu = \mathbf{E}[s(\boldsymbol{y})]$. Let C be the convex hull of the set $\{s(\boldsymbol{y}) : \boldsymbol{y} \in \mathcal{Y}\}$, $ri(C)$ its relative interior and $rbd(C)$ its relative boundary. It turns out that if $\mu(\boldsymbol{\theta})$ is close to $rbd(C)$, that the model places most of the probability mass on graphs belonging to the set $deg(\mathcal{Y}) = \{\boldsymbol{y} \in \mathcal{Y} : s(\boldsymbol{y}) \in rbd(C)\}$. It is also known that the MLE exists if and only if $s(\boldsymbol{y}) \in ri(C)$ and if it exists it is unique. In the context of the likelihood approximations which we present in the next Section, it will be important to examine whether the corresponding approximate posterior distribution supports parameter values in the degenerate region.

3   Likelihood approximations

In this section we introduce three likelihood approximations for the p^* model. Suppose that the collection of all possible dyads have been ordered as $(y_1, y_2, \ldots, y_n)$. From this point onward, for ease of notation, we will denote each dyad by a single index. We will also use the notation $y_{1:i}$ to denote the edges $\{y_1, \ldots, y_i\}$ and y_{-i} to denote the edges $\{y_{1:i-1}, y_{i+1:n}\}$.

3.1   Maximum pseudolikelihood estimation (MPLE)

A standard approach to approximate the distribution of a Markov random field is to use a pseudolikelihood approximation, first proposed in Besag (1974) and adapted for social network models in Strauss and Ikeda (1990). This approximation consists of a product of easily normalised full-conditional distributions

$$
\begin{aligned}
\pi(\boldsymbol{y}|\beta) &\approx \pi_{pseudo}(\boldsymbol{y}|\beta) \\
&= \prod_{i=1}^{n} \pi(y_i|\boldsymbol{y}_{-i}, \beta) \\
&= \prod_{i=1}^{n} \frac{\pi(y_i = 1|\boldsymbol{y}_{-i}, \beta)^{y_i}}{\left[1 - \pi(y_i = 0|\boldsymbol{y}_{-i}, \beta)\right]^{y_i-1}}. \tag{2}
\end{aligned}
$$

The basic idea underlying this method is the assumption of weak dependence between the variables in the graph so that the likelihood can be well approximated by the pseudolikelihood function. This leads to a fast estimation, and can be implemented using standard generalised linear model software. Nevertheless this approach turns out to be generally inadequate since

it only uses local information whereas the structure of the graph is affected by global interaction.

3.2 Maximum block-pseudolikelihood estimation (MBPLE)

An obvious extension to the pseudolikelihood estimator is the block-pseudolikelihood estimator. The MBPLE relies on the same idea, but evaluates the full conditional of blocks of variables rather than the full conditionals of single variables.

$$\pi(\boldsymbol{y}|\beta) \approx \pi_{blockpseudo}(\boldsymbol{y}|\beta) = \prod_{b=1}^{B} \pi(\boldsymbol{y}_b|\boldsymbol{y}_{-b},\beta), \quad (3)$$

for some partitioning of the dyads into B disjoint groups of size b, such that $\bigcup_{b=1}^{B} y_b = \boldsymbol{y}$ and $\bigcap_{b=1}^{B} y_b = \varnothing$. Assuming the largest of the blocks contains no more than roughly 20 dyads we calculate each full conditional. This approximation is also quite fast, and attempts to capture larger interactions within the graph. Note that similar block pseudolikelihood approximations have been considered in the context of hidden binary Markov random fields, see Rydén and Titterington (1998) and Friel et al. (2009), where superior performance of the block pseudolikelihood estimators was observed.

3.3 Relaxed dependence approximation (RDA)

The joint distribution of $\boldsymbol{y}$ can be written as

$$p(\boldsymbol{y}|\beta) = p(y_{1:b}|y_{b+1:n},\beta) \prod_{i=b+1}^{n-1} p(y_i|y_{i+1:n},\beta)p(y_n|\beta).$$

The RDA, in essence, attempts to approximate the distribution of $p(y_i|y_{i+1:n},\beta)$. First, notice that

$$\begin{aligned} p(y_i|y_{i+1:n},\beta) &= \frac{p(y_{1:i-1},y_i|y_{i+1:n},\beta)}{p(y_{1:i-1}|y_{i:n},\beta)} \\ &= \frac{p(y_{\bar{A}^i}|y_{i+1:n},\beta)}{p(y_{\bar{A}^i}|y_{i:n},\beta)} \times \\ &\quad \frac{p(y_{A^i},y_i|y_{-\{A^i,i\}},\beta)}{p(y_{A^i}|y_{-\{A^i\}},\beta)} \end{aligned} \quad (4)$$

where $A^i = \{A_1^i,\ldots,A_{b-1}^i\} \subset \{1,\ldots,i-1\}$, $\bar{A}^i = \{1,\ldots,i-1\} \setminus A^i$ and $y_{-\{A\}} = \boldsymbol{y} \setminus y_A$, for $i = b+1,\ldots,n-1$. We introduce an approximation to (4) by writing

$$p(y_i|y_{i+1:n},\beta) \approx \frac{p(y_{A^i},y_i|y_{-\{A^i,i\}},\beta)}{p(y_{A^i}|y_{-\{A^i\}},\beta)} \quad (5)$$

Similarly, we approximate

$$p(y_n) \approx \frac{p(y_{A^n},y_n|y_{-\{A^n,n\}},\beta)}{p(y_{A^n}|y_{-\{A^n\}},\beta)}. \quad (6)$$

Here we define a *block* of size b to be the set $\{y_{A^i},y_i\}$, for $i = b+1,\ldots,n$. We further denote a block of size 1 to correspond to $A^i = \emptyset$, and in this instance, (5) and (6) reduce to

$$p(y_i|y_{i+1:n},\beta) \approx p(y_i|y_{-\{i\}},\beta)$$

and

$$p(y_n|\beta) \approx p(y_n|y_{-\{n\}},\beta),$$

respectively.

Plugging (5) and (6) into (4) yields the approximation

$$\begin{aligned} p(y_1,\ldots,y_n|\beta) &\approx p(y_{1:b}|y_{b+1:n},\beta) \\ &\prod_{i=b+1}^{n-1} \frac{p(y_{A^i},y_i|y_{-\{A^i,i\}},\beta)}{p(y_{A^i}|y_{-\{A^i\}},\beta)} \\ &\times \frac{p(y_{A^n},y_n|y_{-\{A^n,n\}},\beta)}{p(y_{A^n}|y_{-\{A^n\}},\beta)}. \end{aligned} \quad (7)$$

In effect, (5) and (6) assume that

$$\frac{p(y_{A^i},y_i|y_{-\{A^i,i\}},\beta)}{p(y_{A^i}|y_{-\{A^i\}},\beta)} = 1, \quad i = b+1,\ldots,n-1 \quad (8)$$

and

$$\frac{p(y_{\bar{A}^n}|\beta)}{p(y_{\bar{A}^n}|y_n,\beta)} = 1, \quad (9)$$

respectively.

Finally, a nice property of our approximation is that it can be seen as a natural expansion of the pseudolikelihood approximation, which corresponds to a block size of 1. Note that an estimator similar to RDA has been explored in the context of binary Markov random fields on the lattice Friel et al. (2009), and has been implemented in a variational Bayes setting in McGrory et al. (2009).

3.3.1 Ordering the dyads and selecting each blocks

The RDA and MBPLE approaches require that the n dyads in $\boldsymbol{y}$ follow some index ordering. Moreover, for the RDA approach, there is a need to choose, for each $i = b+1,\ldots,n$, the set $y_{A_i} \subset \{y_1,\ldots,y_{i-1}\}$ of dyads in block $\{y_{A_i},y_i\}$. It is unclear to us how to provide guidance for the former requirement. We are able to offer some guidance as to how the set of dyads y_{A_i} is chosen, however. Our intuition is that each block should consist of as many dyads from $\{y_1,\ldots,y_{i-1}\}$ which share a common node with y_i, since these are the dyads which most influence y_i. If there are more than $b-1$ dyads in $\{y_1,\ldots,y_{i-1}\}$ sharing a node with y_i, then $b-1$ such dyads are chosen uniformly at random. While if there are less than $b-1$ dyads sharing a common node with y_i, each of these are selected, and the remainder chosen uniformly at random from the set of dyads not sharing a common node with y_i.

4 Monte Carlo approaches

An alternative to approximating the likelihood is to try to estimate the true posterior distribution. This is the approach taken by the Monte Carlo maximum likelihood (MC-MLE) algorithm introduced by Geyer and Thompson (1992). This algorithm has been widely used to carry out maximum likelihood estimation for the p^* model. A key identity is the following

$$\frac{z(\beta)}{z(\beta_0)} = \mathbf{E}_{\boldsymbol{y}|\beta_0} \left[\frac{q(\boldsymbol{y}|\beta)}{q(\boldsymbol{y}|\beta_0)} \right]$$
$$= \sum_{\boldsymbol{y}} \frac{q(\boldsymbol{y}|\beta)}{q(\boldsymbol{y}|\beta_0)} \frac{q(\boldsymbol{y}|\beta_0)}{z(\beta_0)}$$
$$\approx \frac{1}{m} \sum_{i=1}^{m} \exp \left\{ (\beta - \beta_0)^t s(\boldsymbol{y}_i) \right\}$$

where β_0 is fixed set of parameter values, and $\mathbf{E}_{\boldsymbol{y}|\beta_0}$ denotes an expectation taken with respect to the distribution $\pi(\boldsymbol{y}|\beta_0)$. In practice this ratio of normalising constants is approximated using graphs $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ sampled via MCMC from the stationary distribution defined by β_0 and importance sampling. This yields the following approximated log likelihood ratio:

$$\hat{l}_{\beta_0}(\beta) \approx l(\beta_0) + (\beta - \beta_0)^t s(\boldsymbol{y}) -$$
$$\log \left[\frac{1}{m} \sum_{i=1}^{m} \exp \left\{ (\beta - \beta_0)^t s(\boldsymbol{y}_i) \right\} \right]. \quad (10)$$

This is then viewed as a function of β, and its maximum value serves as a Monte Carlo estimate of the MLE.

A crucial aspect of this algorithm is the choice of β_0. Ideally β_0 should be very close to the maximum likelihood estimator of β. In fact $\hat{l}_{\beta_0}(\beta)$ is very sensitive to the choice of β_0. A poorly chosen value of β_0 may lead to an objective function (10) that cannot even be maximised, see Handcock (2003).

In pratice, β_0 is often chosen as the maximiser of (2), although this itself maybe a very biased estimator. Indeed, (10) may also be sensitive to numerical instability, since it effectively computes the ratio of a normalising constant, but it is well understood that the normalising constants can vary by orders of magnitude with $\boldsymbol{\theta}$.

5 Examples

In this section we consider two real dataset to test and compare the different methods introduced in the previous section. For each example we choose to fit the same model as was introduced in section 2. For both datasets our goal is to gather information about the posterior distribution of β, using our likelihood approximations, $\tilde{p}(\boldsymbol{y}|\beta)$,

$$p(\beta|\boldsymbol{y}) \propto p(\boldsymbol{y}|\beta)p(\beta) \approx \tilde{p}(\boldsymbol{y}|\beta)p(\beta). \quad (11)$$

We do this by proceeding in the following manner. First we locate the mode of the posterior distribution by plugging our approximate likelihood function combined with an uninformative prior into a black-box optimiser. Once the mode is located we design a grid surrounding the mode and evaluate an approximate unnormalised posterior distribution (the right hand side of (11)) at each grid point. This allows us to produce numerical approximations to marginals, means, variances and other aspects of the posterior distribution. We also want to ensure that our approximations have not moved us into parameter space that produces degenerate graphs. So once the posterior is calculated we sample 500 values for β from the posterior. We then use the `ergm` package for R (Hunter et al., 2008) to simulate a graph from each of these sets of parameters and study the resulting graphs to check for degeneracy. For both the RDA and blockpseudo the blocksize is critical, the bigger the blocks, the better we expect our approximation to be, but at the price of increased run-time, also, avaliable memory restricts us to blocks of size ≤ 20. As mentioned earlier, a block size of 1 is equivalent to standard pseudolikelihood.

5.1 Molecule example

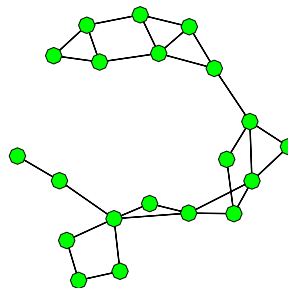Our first example examined the dataset illustrated in figure 1.

Figure 1: Molecule graph.

The graph consists of 20 nodes in a quite sparse configuration, which gives us a dataset of 190 variables. In figure 2 we have plotted the approximate posterior mode returned from the optimisation algorithm for the RDA and blockpseudo approach, with blocksizes ranging from 1 to 18 for the rda and 1 to 16 for blockpseudo.
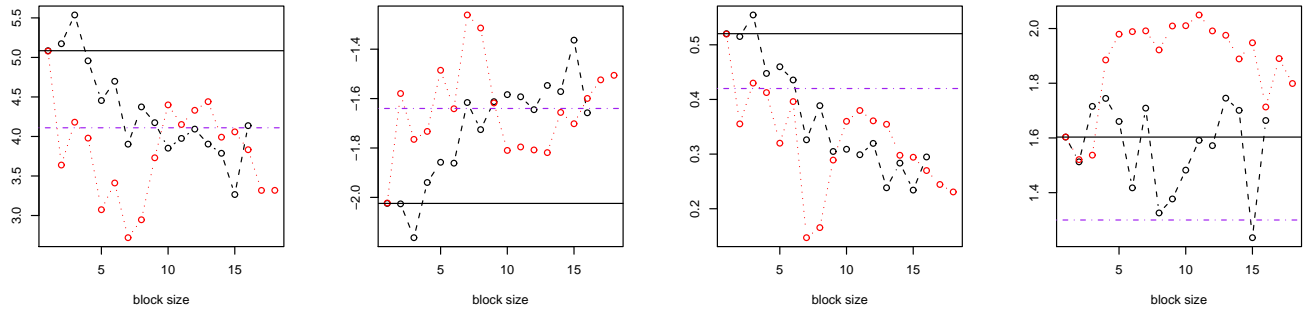
Figure 2: Molecule dataset: Approximate maximum a posteriori parameters estimates for β_1, β_2, β_3 and β_4, plotted from left to right for blocksizes from 1 to 16 (blockpseudo) and 1 to 18 (rda). The black dotted line represents the blockpseudo approach while the red stapled line represents the RDA approximation. The horizontal black line represents the pseudolikelihood approximation, while the horizontal purple stapled line represents an MC-MLE approximation method.
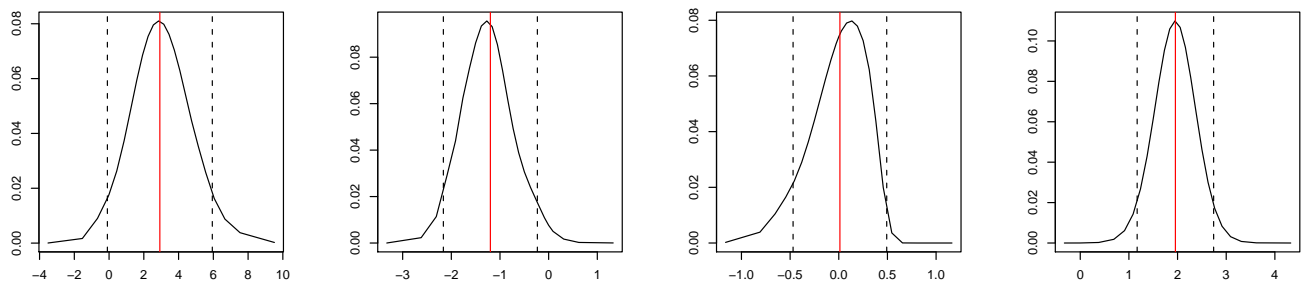

Figure 3: Molecule dataset: Approximate posterior marginals for β_1, β_2, β_3 and β_4 using the RDA approximation. The red dotted line represents the mean and the black dotted lines indicate distances of two posterior standard deviations from the mean.

The vertical line is placed at the value for blocksize 1, which represents the pseudolikelihood approximation. As we can see the RDA and blockpseudo seem to give similar parameter estimates (with the exception of β_4) and each parameter estimate is quite different from the pseudolikelihood estimator. The purple stapled line represents an estimate returned by an MC-MLE method. We note that this comes closer to the approximations returned by the RDA and blockpseudo. The posterior distribution of β was next evaluated in a grid surrounding the mode of the RDA approximation with a blocksize of 10. The grid contained 25 points in each dimension. Figure 3 shows the approximate marginal distributions $p(\beta_i|\boldsymbol{y})$ with means and credible intervals, calculated from the full posterior distribution. Note that parameter estimates for pseudolikelihood and MC-MLE found using the `ergm` package gave rise to realisations that are almost all degenerate. By constrast realisations conditional on parameters from the posterior were not degenerate, being neither complete or empty.

We also estimate the marginal distribution of the data. Recall that in equation (11) the normalizing constant is the approximate marginal likelihood of y. Hence we can estimate the marginal likelihood of our data by completing the finite sum over the parameters,

$$p(\boldsymbol{y}) \approx \tilde{p}(\boldsymbol{y}) = \sum_{\beta_1} \sum_{\beta_2} \sum_{\beta_3} \sum_{\beta_4} \tilde{p}(\boldsymbol{y}|\beta)p(\beta). \qquad (12)$$

Here, recall that the summands on the right hand side of (12) are available and therefore an estimate of the marginal likelihood results from summing these over all grid points, β. The grid computed for the saturated model can then be used to estimate the marginal likelihood for any model nested within the saturated model, by simply setting those parameter values which are not included in the model to zero. This was then done for all 15 different model configurations containing at least one parameter. Assigning equal weights to each of the models the posterior model probabilities can be estimated as,

$$\tilde{p}(m_i|y) = \frac{\tilde{p}(y|m_i)p(m_i)}{\sum_j \tilde{p}(y|m_j)p(m_j)} \ i = 1, \ldots, 15. \qquad (13)$$

The models with highest posterior model probability turned out to be the saturated model, the model containing the parameters $\{\beta_2, \beta_3, \beta_4\}$ and the model containing the parameters $\{\beta_1, \beta_2, \beta_4\}$, these achieved probabilities 0.64, 0.14 and 0.11 respectively.

5.2 Karate example

For our second example we studied the dataset illustrated in figure 4.
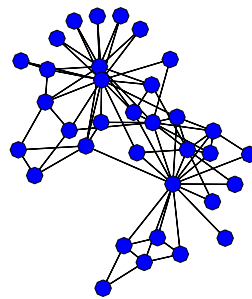

Figure 4: Karate graph.

This graph is larger than the previous example and consists of 34 nodes in a not so sparse configuration, which gives us a dataset of 561 variables. In figure 5 we have plotted the approximate posterior mode returned from the optimisation algorithm for the RDA and blockpseudo approach, with blocksizes ranging from 1 to 16. As in the plot in the previous example the vertical line is placed at the value for blocksize 1, which represents the pseudolikelihood approximation. The RDA and blockpseudo seem to return slightly, but not entirely dissimilar parameter estimates and again each parameter estimate is quite different from the pseudolikelihood estimator. We evaluated the posterior distribution of β as in the previous example with a grid surrounding the mode of the RDA approximation with a blocksize of 10. The grid contained 20 points in each dimension. Figure 6 shows the approximate marginal distributions $p(\beta_i|\boldsymbol{y})$ with means and credible intervals, calculated from the full posterior distribution. As for the previous example, parameter estimates for pseudolikelihood and MC-MLE found using the `ergm` package gave rise to realisations that are almost all degenerate. By constrast realisations conditional on parameters from the posterior based on RDA and block pseudolikelihood were not degenerate, being neither complete or empty. Exactly as in the molecule example we also estimate the marginal distribution of the data. The two models with highest posterior model probability turned out to be the saturated model and the model containing the parameters $\{\beta_1, \beta_2, \beta_3\}$, these achieved probabilities 0.705 and 0.236 respectively.

6 Discussion

Despite the widespread use of the p^* model in social network analysis, the inferential methods used to service this model are lacking in many respects. The approximations which we have outlined in this paper ad-
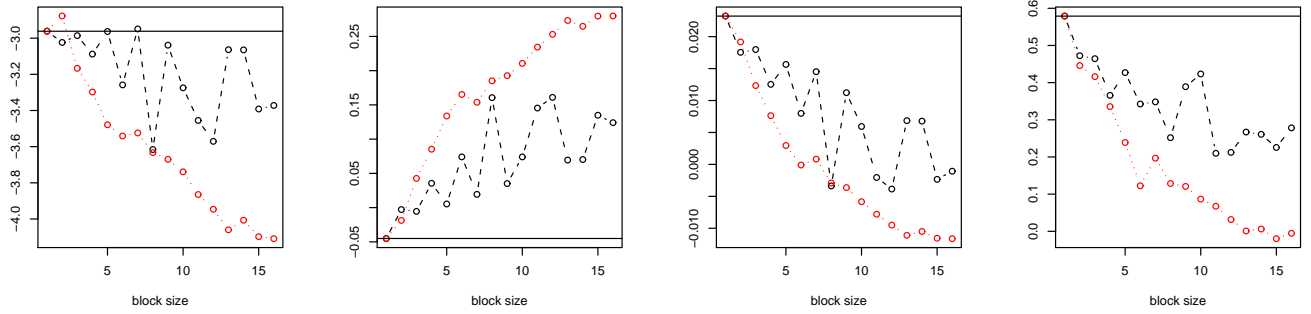
Figure 5: Karate dataset: Approximate maximum a posteriori parameters estimates for β_1, β_2, β_3 and β_4, plotted from left to right for blocksizes from 1 to 16. The black dotted line represents the blockpseudo approach while the red stapled line represents the RDA approximation. The horizontal black line represents the pseudolikelihood approximation, while the horizontal purple stapled line represents an MC-MLE approximation method.
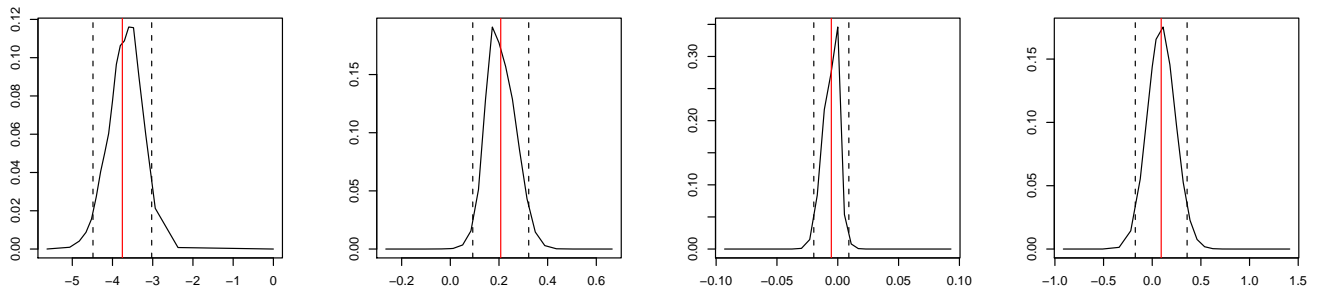

Figure 6: Karate dataset: Approximate posterior marginals for β_1, β_2, β_3 and β_4 using the RDA approximation. The red dotted line represents the mean and the black dotted lines indicate distances of two posterior standard deviations from the mean.

dress this issue. Both of these approximations extend the standard pseudolikelihood approximation. Our approximations can be considered as composite likelihood approximations, where the composite factors in the likelihood involve at most 20 dyad variables. Composite likelihood methods are popular in the statistics literature, for example Heagerty and Lele (1998), Cox and Reid (2004), and the theory surrounding such methods is well established. However composite likelihoods have received relatively little attention in the Bayesian literature, and future work in this directions would be useful.

As mentioned in section 3, it is unclear how to choose an index ordering of the dyads for RDA and blockpseudolikelihood. However for the RDA approach, we believe that including as many as of the dyads which share a common node with y_i, represents a reasonable way to select composite blocks.

We note that our inference methods for the p^* model provide an appealing simulation-free alternative to the usual Markov chain Monte Carlo approaches. In particular our methods can be considered as an inference machine for these types of model providing the end user with the possibility to explore probabilistic uncertainty for the model parameters and also for the uncertainty estimates for the model itself. Finally, we are currently automating our computer code to provide the end user with a suite of routines to carry out the inference tasks outlined in this paper.

References

Besag, J. E. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

Cox, D. R. and Reid, N. (2004), "A note on pseudolikelihood constructed from marginal densities," *Biometrika*, 91, 729–737.

Erdös, P. and Rényi, A. (1959), "On random graphs," *Publicationes Mathematicae*, 6, 290–297.

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

Friel, N., Pettitt, A. N., Reeves, R., and Wit, E. (2009), "Bayesian inference in hidden Markov random fields for binary data defined on large lattices," *Journal of Computational and Graphical Statistics*, 18, 243–261.

Geyer, C. J. and Thompson, E. A. (1992), "Constrained Monte Carlo maximum likelihood for dependent data (with discussion)," *Journal of the Royal Statistical Society, Series B*, 54, 657–699.

Handcock, M. S. (2003), "Assessing Degeneracy in Statistical Models of Social Networks," *Working Paper no.39, Center for Statistics and the Social Sciences, University of Washington.*

Heagerty, P. J. and Lele, S. R. (1998), "A composite likelihood approach to binary spatial data," *Journal of the American Statistical Association*, 93, 1099–1111.

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., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008), "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks," *Journal of Statistical Software*, 24, 1–29.

McGrory, C. A., Titterington, D. M., Reeves, R., and Pettitt, A. N. (2009), "Variational Bayes for estimating the parameters of a hidden Potts model." *Statistics and Computing*, 19, 329–340.

Rinaldo, A., Fienberg, S. E., and Zhou, Y. (2009), "On the geometry of descrete exponential random families with application to exponential random graph models," *Electronic Journal of Statistics*, 3, 446–484.

Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007), "Recent developments in exponential random graph (p^*) models for social networks," *Social Networks*, 29, 192–215.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion)," *Journal of the Royal Statistical Society, Series B*, 71, 319–392.

Rydén, T. and Titterington, D. M. (1998), "Computational Bayesian analysis of hidden Markov models," *Journal of Computational and Graphical Statistics*, 7, 194–211.

Strauss, D. and Ikeda, M. (1990), "Pseudolikelihood estimation for social networks," *Journal of the American Statistical Association*, 5, 204–212.

Wasserman, S. and Pattison, P. (1996), "Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^*," *Psycometrica*, 61, 401–425.