*Appl. Statist.* (2004)
**53**, *Part* 2, *pp.* 249–260

# Bayesian analysis of directed graphs data with applications to social networks

Paramjit S. Gill

*Okanagan University College, Kelowna, Canada*

and Tim B. Swartz

*Simon Fraser University, Burnaby, Canada*

[Received September 2002. Revised July 2003]

**Summary.** A fully Bayesian analysis of directed graphs, with particular emphasis on applications in social networks, is explored. The model is capable of incorporating the effects of covariates, within and between block ties and multiple responses. Inference is straightforward by using software that is based on Markov chain Monte Carlo methods. Examples are provided which highlight the variety of data sets that can be entertained and the ease with which they can be analysed.

*Keywords*: Bayesian analysis; Markov chain Monte Carlo methods; Social network models; Statistical graph theory; WinBUGS

## 1. Introduction

The analysis of social networks is an important activity that is of interest to many researchers, including sociologists, anthropologists, communication scientists and social psychologists. For example, in organized settings (a business firm, say), the co-operation and exchange of goodwill between members may be important for the well-being of the organization. Each node in the network plays the dual role of an 'actor' (rater or responder) and a 'partner' (target or stimulus). The nodes themselves, or possibly a third-party observer, supplies dyadic data concerning the pair of nodes. Relational ties between nodes are channels for the transfer of resources such as goods, money, information, political support and friendship. In the case where the relationship is binary (yes or no), a social network can be described simply as a directed graph or a digraph. For example, in a study of friendship patterns, a directed edge from individual $i$ to individual $j$ means that individual $i$ says that individual $j$ is a friend. More generally, we use the terminology that $i$ has established a 'tie' with $j$. Note that the individual $j$ in return may or may not consider $i$ as a friend.

A key feature of network data is that the classical assumption of the independence of ties is violated. In fact, the modelling of interdependence is of primary interest. Since the 1930s, a variety of deterministic (graph theoretic) and statistical methods have been used for the analysis of social networks (see, for example, Wasserman and Faust (1994)). In a seminal paper, Holland and Leinhardt (1981) proposed the $p_1$-model for the analysis of directed graphs. The model includes fixed effects parameters for describing the abilities of an individual node to

*Address for correspondence*: Tim B. Swartz, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada.
E-mail: tim@stat.sfu.ca

 0035–9254/04/53249

attract and to produce ties and the tendency of a pair of nodes to reciprocate ties. Wong (1987) investigated an empirical Bayesian approach as an alternative to the fixed effects model. He also presented an EM–Newton type of algorithm for maximum likelihood (ML) estimation of variance–covariance components. The $p_1$-model assumes that the dyads are independent, i.e. the probability of ties between two nodes $i$ and $j$ is not affected by the presence (or absence) of ties involving any other pair of nodes. This is a very strong assumption which is often violated. For example, in a friendship network, if A and B are friends, and A and C are friends, it is then more likely that B and C are friends.

Some attempts have been made to relax the assumption of dyadic independence. Strauss and Ikeda (1990) introduced the use of a pseudolikelihood function (Besag, 1975) which incorporates dyad dependence through modelling the probability of a tie conditionally on the rest of the data. Wasserman and Pattison (1996) and Anderson *et al.* (1999) popularized the use of pseudolikelihood estimation and labelled the associated model as the $p^*$-model. The $p^*$-model allows some degree of dependence through the consideration of network properties such as transitivity, cyclicity and 2-stars (see Wasserman and Pattison (1996) for definitions of these terms). The main motivation for using the pseudolikelihood approach is computational ease. Estimation is carried out by using the standard logistic regression model as implemented in popular statistical software packages and inference is therefore based on asymptotic theory. Besag (2000) contended that the inferential approach in $p^*$-models may be unsuitable. He argued that the number of parameters increases with the sample size and therefore that the approach is unlikely to perform satisfactorily unless the interactions in the system are weak. He suggested the use of ML estimation using Markov chain Monte Carlo (MCMC) techniques. Snijders (2002) reported on issues concerning MCMC-based ML estimation using various data sets; he occasionally encountered difficulties of convergence and large discrepancies when compared with pseudolikelihood estimation.

In this paper we explore the use of random-effects models to incorporate dependence between the dyads. The basic idea was proposed by Wong (1987) and is a natural generalization of the fixed effects $p_1$-model. The node-specific effects model expresses the likelihood of ties in terms of the nodal attributes rather than in terms of network structural properties (such as transitivity and cyclicity). We propose a fully Bayesian approach to network modelling using the $p_1$-model as a basis. MCMC simulation permits the investigation of any posterior characteristic of interest such as marginal posterior distributions. It allows for complicated modelling and produces exact inference regardless of the sample size. This is in contrast with analyses which focus only on the estimation of primary parameters and their asymptotic standard errors. Because of computational limitations, a fully Bayesian approach for network analysis seemed impossible until the 1990s. But, with the advent of modern computational machinery and the development of simulation algorithms, many sophisticated models can now be fitted with relative ease by using the software package WinBUGS (Spiegelhalter *et al.*, 2000). The software allows even unsophisticated users to entertain complex Bayesian models. Gill and Swartz (2001, 2003) used MCMC methods in the Bayesian analysis of round robin interaction data where the response variable is continuous. Nowicki and Snijders (2001) considered MCMC Bayesian analysis for block model structures where the relationship between two nodes depends only on block membership. The inferential emphasis, however, concerns the latent variables that determine the block structure.

In Section 2, we present the $p_1$-model of Holland and Leinhardt (1981) and its basic Bayesian counterpart and we demonstrate how the basic Bayesian model can be modified to cater to special settings. We show how covariates can be introduced and how the stochastic block models (Wang and Wong, 1987) are variations of the basic Bayesian model. We also describe how output from MCMC simulations can be used in model selection involving competing Bayesian

$p_1$-models. In Section 3, detailed analyses of Bayesian $p_1$-models are provided on a well-known data example. In Section 4, some progress is made on the more complex multigraph models which involve multiple-response data. The associated WinBUGS code for Sections 3 and 4 is available from the authors on request.

## 2. Bayesian $p_1$-models

### 2.1. Preliminaries

In social networks studies, $m$ nodes indicate their relationship with one another. Node $i$ indicates its 'relationship' with node $j$ $(\neq i)$ through a binary outcome $X_{ij}=1$ if node $i$ relates to node $j$, and $X_{ij}=0$ otherwise. The $p_1$-model of Holland and Leinhardt (1981) considers the joint distribution of the dyads $D_{ij}=(X_{ij}, X_{ji})$ with dyadic probabilities

$$\begin{aligned} a_{ij} &= \Pr(\text{mutual dyad}) = \Pr\{D_{ij}=(1,1)\}, \\ b_{ij} &= \Pr(\text{asymmetric dyad}) = \Pr\{D_{ij}=(1,0)\}, \\ c_{ij} &= \Pr(\text{null dyad}) = \Pr\{D_{ij}=(0,0)\} \end{aligned}$$

where $i \neq j$ and $a_{ij}+b_{ij}+b_{ji}+c_{ij}=1$. Assuming that the dyads are mutually independent, the joint probability for the data matrix $\mathbf{X}=(X_{ij})$ is

$$\Pr(\mathbf{X}) \propto \exp\Big(\sum_{i<j} \phi_{ij} X_{ij} X_{ji} + \sum_{i\neq j} \theta_{ij} X_{ij}\Big) \tag{1}$$

where

$$\begin{aligned} \phi_{ij} &= \log\left(\frac{a_{ij}c_{ij}}{b_{ij}b_{ji}}\right) = \log\left\{\frac{\Pr(X_{ij}=1|X_{ji}=1)\Pr(X_{ij}=0|X_{ji}=0)}{\Pr(X_{ij}=0|X_{ji}=1)\Pr(X_{ij}=1|X_{ji}=0)}\right\} \\ \theta_{ij} &= \log\left(\frac{b_{ij}}{c_{ij}}\right) = \log\left\{\frac{\Pr(X_{ij}=1|X_{ji}=0)}{\Pr(X_{ij}=0|X_{ji}=0)}\right\}. \end{aligned} \tag{2}$$

In this model, $\phi_{ij}$ and $\theta_{ij}$ are free parameters. The log-odds-ratio $\phi_{ij}$ $(=\phi_{ji})$ measures the degree of reciprocity or mutuality of ties between subjects $i$ and $j$. It is a composite measure whereby larger values of $\phi_{ij}$ indicate that subject $i$ is more likely to establish the same type of relationship with subject $j$ as $j$ has with $i$. The log-odds-ratio $\theta_{ij}$ measures the likelihood of a tie from node $i$ to node $j$ given that $X_{ji}=0$. The original Holland and Leinhardt (1981) model postulates that

$$\begin{aligned} \phi_{ij} &= \phi, \\ \theta_{ij} &= \theta + \alpha_i + \beta_j. \end{aligned} \tag{3}$$

Parameters $\phi$, $\theta$, $\alpha_i$ and $\beta_j$ are of primary interest. The parameter $\theta$ is called the 'density' or choice parameter since it measures the overall degree of forming ties in the population. The parameters $\alpha_i$ and $\beta_i$ represent respectively the properties of expansiveness (the ability to produce ties) and attractiveness (the ability to attract ties) of subject $i$. A positive value of $\alpha_i$ or $\beta_i$ indicates that subject $i$ is more expansive or attractive respectively than a typical member of the population. Holland and Leinhardt (1981) discussed ML estimation of the model parameters. The ML formulation is based on the joint probability distribution which assumes that the dyads $D_{ij}$ are statistically independent. The numerical computation in ML estimation often results in problems of convergence for networks of realistic size.

## 2.2. Bayesian modelling

Wong (1987) proposed a Bayesian approach assuming a random nodes effects model. Conditionally on the node-specific random effects, the dyads are assumed mutually independent. The random-effects model postulates that

$$\left.\begin{aligned} \phi &\sim \text{normal}(\mu_\phi, \sigma_\phi^2), \\ \theta &\sim \text{normal}(\mu_\theta, \sigma_\theta^2), \\ \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} &\sim \text{normal}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\alpha\beta} \right\}, \\ \Sigma_{\alpha\beta} &= \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \end{aligned}\right\} \qquad (4)$$

where $\phi$, $\theta$ and $(\alpha_i, \beta_i)$ are independent.

In the random-effects model, the variance–covariance parameters $\sigma_\phi^2$, $\sigma_\theta^2$, $\sigma_\alpha^2$, $\sigma_\beta^2$ and $\rho$ are of primary interest. The correlation $\rho$ takes into account the association between expansiveness and attractiveness of a subject. A positive value of $\rho$ indicates that 'expansive' subjects are those who are relatively more capable of attracting relational ties. Wong (1987) applied the EM algorithm for obtaining the ML estimates of the variance–covariance matrix and used empirical Bayes estimation for the individual effects $\alpha$s and $\beta$s.

A Gibbs sampling approach is based on generating variates from conditional distributions. To facilitate the use of Gibbs sampling for our purpose, we use an equivalent log-linear formulation as suggested by Fienberg and Wasserman (1981). In this formulation, a dyad $(X_{ij}, X_{ji})$ is represented by four Bernoulli variables $Y_{ij00}, Y_{ij10}, Y_{ij01}$ and $Y_{ij11}$ as follows (see also Wasserman and Faust (1994), chapter 15):

$$Y_{ijkl} = \begin{cases} 1 & \text{if } X_{ij} = k, X_{ji} = l, \\ 0 & \text{otherwise.} \end{cases}$$

The $p_1$-model is then described by the four log-linear equations

$$\begin{aligned} \log\{\Pr(Y_{ij00} = 1)\} &= \lambda_{ij}, \\ \log\{\Pr(Y_{ij10} = 1)\} &= \lambda_{ij} + \theta + \alpha_i + \beta_j, \\ \log\{\Pr(Y_{ij01} = 1)\} &= \lambda_{ij} + \theta + \alpha_j + \beta_i, \\ \log\{\Pr(Y_{ij11} = 1)\} &= \lambda_{ij} + 2\theta + \alpha_i + \alpha_j + \beta_i + \beta_j + \phi \end{aligned}$$

where the joint distribution only includes the non-redundant dyads corresponding to $i < j$. In this formulation, the parameters $\lambda_{ij} = \log(c_{ij})$ are fixed according to the constraint $\Sigma_{k,l} Y_{ijkl} = 1$.

In addition to the specification of prior probability distributions for $\phi$, $\theta$, $\alpha_i$ and $\beta_i$ as in expression (4) above, the Bayesian approach needs a prior specification for the second-stage parameters $\mu_\phi$, $\mu_\theta$, $\sigma_\phi^2$, $\sigma_\theta^2$, $\sigma_\alpha^2$, $\sigma_\beta^2$ and $\rho$. We let

$$\left.\begin{aligned} \mu_\phi &\sim \text{normal}(\mu_0, \sigma_0^2), \\ \mu_\theta &\sim \text{normal}(\mu_0, \sigma_0^2), \\ \sigma_\phi^{-2} &\sim \text{gamma}(a_0, b_0), \\ \sigma_\theta^{-2} &\sim \text{gamma}(a_0, b_0), \\ \sigma_\alpha^{-2} &\sim \text{gamma}(a_0, b_0), \\ \sigma_\beta^{-2} &\sim \text{gamma}(a_0, b_0), \\ \rho &\sim \text{uniform}(-1, 1) \end{aligned}\right\} \qquad (5)$$

where prior independence is assumed.

The parameters that are subscripted with a 0 in expression (5) are referred to as hyperparameters and are often set to give diffuse or non-informative prior distributions for the second-stage parameters. Diffuse distributions are useful when we do not have strong prior opinions regarding parameters. In the examples in this paper, we have used $\mu_0 = 0$, $\sigma_0 = 10000$, $a_0 = 0.0001$ and $b_0 = 0.0001$. Users may consider modifications to the prior specification that is given in expression (5). For example, in principle, the prior mean of $\mu_\phi$ can differ from the prior mean of $\mu_\theta$.

Let $[A|B]$ denote the conditional density of $A$ given $B$. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and let $\beta = (\beta_1, \beta_2, \ldots, \beta_m)$. Having specified the model, the model assumptions and data induce the posterior distribution given by

$$[\phi, \theta, \alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2, \rho, \mu_\phi, \mu_\theta, \sigma_\phi^2, \sigma_\theta^2 | \mathbf{Y}] \propto [\mathbf{Y} | \phi, \theta, \alpha, \beta][\phi | \mu_\phi, \sigma_\phi^2][\theta | \mu_\theta, \sigma_\theta^2] \times [\alpha, \beta | \sigma_\alpha^2, \sigma_\beta^2, \rho][\sigma_\alpha^2, \sigma_\beta^2, \rho][\sigma_\phi^2][\sigma_\theta^2]. \quad (6)$$

The posterior distribution is the probability distribution of the parameters conditional on the data and is the fixture from which inference proceeds. The posterior distribution is complex and difficult to study in its full functional form. In the present case, the posterior distribution is $2m + 8$ dimensional. Typically, however, we are interested solely in the average value and spread of some of the parameters. Therefore, if we can repeatedly generate values of a parameter, say $\sigma_\alpha^2$, from the posterior distribution, average those values and calculate their sample standard deviation, we shall then have obtained estimates of the desired posterior mean and posterior standard deviation.

Generating variates directly from multivariate posterior distributions is typically difficult. A popular method is the generation of variates from a simpler Markov chain for which the posterior distribution is the equilibrium distribution. This general approach is known as MCMC sampling and it has revolutionized the practice of Bayesian statistics. However, the implementation of MCMC sampling in Bayesian applications is not always straightforward. The user first needs to determine a particular Markov chain which has the posterior distribution as the equilibrium distribution. Second, the user may need programming expertise as it is often necessary to generate variables from non-standard distributions. Fortunately, WinBUGS is a high level software package that often avoids these difficulties and allows a user to entertain complex Bayesian models. The user needs only to supply the data and to provide the model specification. Details relating to the Markov chain are performed in the background. The user deals only with the output from the Markov chain after making sure that the chain has achieved practical convergence. WinBUGS has built in graphical and analytical capabilities to check convergence. We highlight the use of WinBUGS in Section 3.

In any Bayesian analysis, it is recommended that the user checks the sensitivity of inferences to the prior specification. For the social networks models that are considered in this paper, there are some easy checks that can be implemented in this direction. For example, it is simply a matter of changing one term in the WinBUGS code to modify univariate normal distributions to univariate Student distributions. Student distributions have long been used in robustness studies as they have longer tails than normal distributions and are more apt at accommodating outliers. We might also modify the prior density $[\sigma_\alpha^2, \sigma_\beta^2, \rho]$ in expression (5) to an inverse Wishart distribution. Of course, it is also possible to modify the user-specified hyperparameters.

### 2.3. *Stochastic block models*

On the basis of the nodal attributes (e.g. sex), the nodes can be divided into groups or blocks. We emphasize that these blocks are determined *a priori* and are not based on the data to be analysed. The constant density and reciprocity model can then be replaced by a differential density and/or differential reciprocity specification with

$$\begin{aligned}\theta_{ij} &= \theta + \delta G_{ij} + \alpha_i + \beta_j, \\ \phi_{ij} &= \phi + \nu G_{ij}\end{aligned} \tag{7}$$

where the indicator variable $G_{ij} = 1$ if nodes $i$ and $j$ belong to the same group, and $G_{ij} = 0$ otherwise. The parameters $\delta$ and $\nu$ measure the effects of same group membership on the density and reciprocity respectively. Of course, there are other possibilities, such as $\delta$ or $\nu$ being different over group pairs (e.g. different for male–male, female–female and male–female pairs). Some prior assumptions also need to be made and we construct them in the same spirit as the prior assumptions in the basic model. For example, we assume $\delta \sim \text{normal}(\mu_\delta, \sigma_\delta^2)$ with $\mu_\delta \sim \text{normal}(\mu_0, \sigma_0^2)$ and $\sigma_\delta^{-2} \sim \text{gamma}(a_0, b_0)$. Similar assumptions are made concerning $\nu$ and its hyperparameters.

A more elaborate extension is the model where all or some of the variance–covariance parameters $\{\sigma_\alpha^2, \sigma_\beta^2, \rho\}$ are group specific. In the context of the friendship network, for example, we may be interested in knowing whether males differ more than do females in their expansiveness (i.e. whether $\sigma_{\mathrm{M}\alpha}^2$ is larger than $\sigma_{\mathrm{F}\alpha}^2$). Likewise, similar hypotheses for the other parameters can be formulated and tested. Naturally, we should be careful about introducing too many additional parameters as this may cause problems of identifiability.

### 2.4. *Model selection*

Currently, there is no consensus on the 'correct' approach to Bayesian model selection; see, for example, chapter 6 of Gelman *et al.* (1995). We present three popular approaches to Bayesian model selection. In theory, each method is simple to use since the relevant quantities can be calculated directly from MCMC output.

The first method of model selection is based on the predictive performance criterion that was proposed by Laud and Ibrahim (1995). Given a finite number of candidate models, the criterion is based on the predictive performance of a model in terms of its ability to predict a replicate of the data. The original Bernoulli outcome $X_{ij}$ has a probability of 'success' in terms of $p_1$-model parameters as follows:

$$\begin{aligned}\pi_{ij} &= \Pr(X_{ij} = 1) \\ &= \Pr(Y_{ij10} = 1) + \Pr(Y_{ij11} = 1) \\ &= \exp(\lambda_{ij} + \theta_{ij}) + \exp(\lambda_{ij} + \theta_{ij} + \theta_{ji} + \phi_{ij}).\end{aligned} \tag{8}$$

In every MCMC loop, a replicate $X_{ij\text{pred}} \sim \text{Bernoulli}(\pi_{ij})$ can be generated conditionally on the currently generated values of the model parameters.

Let $\mathbf{X}_{\text{pred}}$ denote a replicate of the observed data matrix $\mathbf{X}_{\text{obs}}$. The posterior predictive distribution of $\mathbf{X}_{\text{pred}}$ under model $M$ is

$$f^{(M)}(\mathbf{X}_{\text{pred}}|\mathbf{X}_{\text{obs}}) = \int f(\mathbf{X}_{\text{pred}}|\eta^{(M)})\, f(\eta^{(M)}|\mathbf{X}_{\text{obs}})\, \mathrm{d}\eta^{(M)} \tag{9}$$

where $\eta^{(M)}$ denotes all the parameters under model $M$, $f(\eta^{(M)}|\mathbf{X}_{\text{obs}})$ is the posterior density and $f(\mathbf{X}_{\text{pred}}|\eta^{(M)})$ is the density of the predicted (or future) value. The model selection criterion, called the expected predictive deviance (EPD), chooses the model $M$ with the smallest value of

$$E^{(M)}\{d(\mathbf{X}_{\text{pred}}, \mathbf{X}_{\text{obs}})|\mathbf{X}_{\text{obs}}\} \tag{10}$$

where $d(\mathbf{X}_{\text{pred}}, \mathbf{X}_{\text{obs}})$ is a discrepancy function and the expectation is with respect to the predictive distribution (9). For binary data, a common discrepancy function is $d(\mathbf{X}_{\text{pred}}, \mathbf{X}_{\text{obs}}) = ||\mathbf{X}_{\text{pred}} - \mathbf{X}_{\text{obs}}||^2$ which is the total number of wrongly predicted ties by model $M$. It is straightforward to estimate expectation (10) as part of MCMC sampling. In each loop of an MCMC run, the matrix $\mathbf{X}_{\text{pred}}$ is generated and $d(\mathbf{X}_{\text{pred}}, \mathbf{X}_{\text{obs}})$ is calculated. The sample mean of the $d(\mathbf{X}_{\text{pred}}, \mathbf{X}_{\text{obs}})$ values is then used to estimate expectation (10).

A related posterior predictive checking approach computes the statistic (Gelman *et al.*, 1996; Daniels and Gatsonis, 1999)

$$\chi^2(\mathbf{X}, \eta^{(M)}) = \sum_i \sum_j \frac{(X_{ij} - \pi_{ij}^{(M)})^2}{\pi_{ij}^{(M)}(1 - \pi_{ij}^{(M)})}$$

where $\pi_{ij}^{(M)}$ is generated via equation (8) through posterior sampling of the parameters under model $M$. As a part of MCMC sampling, a $p$-value can be computed as the proportion of times that $\chi^2(\mathbf{X}_{\text{pred}}, \eta^{(M)})$ exceeds $\chi^2(\mathbf{X}_{\text{obs}}, \eta^{(M)})$. This $p$-value measures the consistency between the variation that is observed in the data and the variation that is predicted by the model. Extreme $p$-values (near 0 or 1) indicate inconsistency in the model (Daniels and Gatsonis, 1999). A $p$-value that is close to 0.5 indicates good agreement between the observed and predicted variations.

The third method of model selection which we consider is based on the more traditional calculation of Bayes factors (Kass and Raftery, 1995). We compare models $M_1$ and $M_2$ by estimating the Bayes factor with

$$\widehat{\text{BF}} = \frac{\sum_k f_1(\mathbf{X}_{\text{obs}} | \eta_k^{(M_1)})}{\sum_k f_2(\mathbf{X}_{\text{obs}} | \eta_k^{(M_2)})}$$

where the $\eta_k$ are output from the Markov chain and $f_i$ denotes the density of the data under model $M_i$. If $\widehat{\text{BF}}$ is greater than or less than 1, this provides evidence in the direction of model $M_1$ or $M_2$ respectively. Unfortunately, a major difficulty with Bayes factors (Kass and Raftery, 1995), and something that we experienced in the following examples, is that the numerator and denominator in $\widehat{\text{BF}}$ are dominated by a few large values. This leads to computational problems and poor estimates of Bayes factors. The Bayes factor approach therefore may only be appropriate for small or moderate-sized social networks.

## 3. Example: corporate law firm data

The data arise from a case-study on a north-eastern US corporate law firm involving 71 lawyers (36 partners and 35 associates). The lawyers were asked to evaluate their colleagues on various relationships. We focus on the directed relationship 'advice' which is an important production-related activity. Subjects were asked to identify colleagues to whom they went for basic professional advice or consultation during the past year. A more comprehensive discussion of the data is provided by Lazega and Pattison (1999). We fit two Bayesian $p_1$-models: one without and the other with the partner–associate effect to investigate how the advice relationship varies with status. We refer to the two models as the base model and the block model respectively. The block model stipulates differential density and differential reciprocity as specified in equations (7) with $G_{ij} = 1$ if lawyers $i$ and $j$ are both partners or both associates, and $G_{ij} = 0$ otherwise. In addition, we also assume block-specific variance–covariance matrices. Summary results from the analyses are provided in Table 1.

**Table 1.** Posterior means and standard deviations SD of some model parameters fitted to the advice response in the corporate law firm data

| *Base model parameters* | | | *Block model parameters*† | | |
|---|---|---|---|---|---|
| *Parameter* | *Mean* | *SD* | *Parameter* | *Mean* | *SD* |
| $\sigma^2_\alpha$ | 1.02 | 0.21 | $\sigma^2_{P\alpha}$ | 0.66 | 0.21 |
| $\sigma^2_\beta$ | 0.69 | 0.15 | $\sigma^2_{A\alpha}$ | 2.58 | 1.00 |
| $\rho$ | −0.28 | 0.13 | $\sigma^2_{P\beta}$ | 1.21 | 0.39 |
| $\theta$ | −2.37 | 0.12 | $\sigma^2_{A\beta}$ | 0.66 | 0.21 |
| $\phi$ | 1.82 | 0.13 | $\rho_P$ | −0.33 | 0.18 |
| | | | $\rho_A$ | −0.41 | 0.19 |
| | | | $\theta$ | −2.60 | 0.21 |
| | | | $\phi$ | 1.74 | 0.26 |
| | | | $\delta$ | 0.94 | 0.10 |
| | | | $\nu$ | 0.02 | 0.27 |

†P denotes partner and A denotes associate.

In the base model, we observe that all the parameters are significant. The values of $\theta$ and $\phi$ suggest that, on average, the lawyers tend to ask advice (and perhaps trust) from those who in turn ask advice of them. We also observe that there is a difference between the lawyers ($\sigma^2_\alpha$ and $\sigma^2_\beta$) in terms of their willingness to ask and be sought for advice. Since $\sigma^2_\beta < \sigma^2_\alpha$, we deduce that there is less variability in lawyers being sought as advisees than in the tendency to seek advice. Interestingly, the negative value of $\rho$ indicates that lawyers who provide much advice are less likely to seek advice.

To assess the sensitivity of the base model to the prior specification, we make several changes to the prior. Specifically, the priors for $\phi$ and $\theta$ in equations (4) and the priors for $\mu_\phi$ and $\mu_\theta$ in equations (5) are modified from normal to Student distributions. We also modify $[\sigma^2_\alpha, \sigma^2_\beta, \rho]$ as specified in equations (5) to an inverse Wishart distribution with an identity scale matrix and 2 degrees of freedom. Under these changes of prior, we observe no meaningful differences in the estimated posterior means and standard deviations. For this data set, the large sample size seems to overcome the prior opinion.

In the block model, the results are similar to the simpler model but with naturally larger standard deviations. However, it appears that the partners behave differently from the associates. For example, we observe that $\sigma^2_{A\alpha} > \sigma^2_{P\alpha}$, which implies that associates are more heterogeneous in seeking advice than the partners. Also $\sigma^2_{P\beta} > \sigma^2_{P\alpha}$ implies that there is more variability in partners being sought as advisees than in the tendency of partners to seek advice. This may be understandable as partners are senior lawyers who are uniformly reluctant to seek advice beyond a small circle of confidants. The positive value of $\delta$ indicates that there is a greater tendency for seeking advice between lawyers of the same professional status. An almost null value of $\nu$ indicates that group membership has no bearing on the degree of reciprocity.

Comparing the two models, we obtain the estimated EPD values of 1229.4 and 1168.3 for the base model and the block model respectively. The posterior predictive checking approach gives $p$-values of 0.46 and 0.51 for the same two models. This suggests that both models are acceptable and we therefore recommend the simpler base model.

To highlight further the ease with which Bayesian models can be fitted and tested by using WinBUGS, we consider the introduction of a continuous covariate to the base model. On the basis of the observation that $\nu$ was not significant in the block model, we consider an alternative model with

$$\begin{aligned}\phi_{ij} &= \phi + \tau w_{ij},\\ \theta_{ij} &= \theta + \alpha_i + \beta_j\end{aligned} \tag{11}$$

where the covariate $w_{ij}$ in equations (11) is the absolute value of the difference in the number of years of working experience between lawyer $i$ and lawyer $j$. Use of the absolute value retains the required symmetry $\phi_{ij} = \phi_{ji}$. A further reason for expanding on $\phi_{ij}$ is that in many data sets $\phi_{ij}$ and $\theta_{ij}$ convey related information. This is apparent from equations (2) whereby $\phi_{ij} = \log(a_{ij}/b_{ji}) - \theta_{ij}$. Therefore it may be sensible to expand the modelling assumptions on $\phi_{ij}$ to be more in line with the extensive modelling assumptions on $\theta_{ij}$. The prior assumptions on $\tau$ are similar to those of $\delta$ in Section 2.3.

The estimates from the analysis of model (11) are similar to the estimates that are obtained in the base model. We observe that $\tau$ is significant with posterior mean −0.09 and posterior standard deviation 0.01. Therefore the effect is such that the lawyers tend to confide more often in colleagues with similar years of experience. The EPD value is 1209.0 and the $p$-value from posterior predictive checking is 0.64. This indicates that model (11) is also an acceptable model.

## 4. Multigraph $p_1$-models

For a detailed analysis of a closed network we should look at the interdependence of various types of relationships. For example, co-operation in an organization can be studied empirically by examining the routine transfers or exchanges between members of various kinds of resources (Lazega and Pattison, 1999). Accordingly, it is desirable to model the information simultaneously as opposed to separate analyses on each response.

In the simplest scenario, consider data arising from two responses. We are faced with two pairs of dyadic data: $D_{ij}^{(1)} = (X_{ij}^{(1)}, X_{ji}^{(1)})$ and $D_{ij}^{(2)} = (X_{ij}^{(2)}, X_{ji}^{(2)})$, $i < j$. Generalizing the discussion in Section 2, it is possible to express the joint probability distribution as

$$\begin{aligned}\Pr(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \propto \exp\Big( &\sum_{i<j} \phi_{ij}^{(1)} X_{ij}^{(1)} X_{ji}^{(1)} + \sum_{i\neq j} \theta_{ij}^{(1)} X_{ij}^{(1)} + \sum_{i<j} \phi_{ij}^{(2)} X_{ij}^{(2)} X_{ji}^{(2)} + \sum_{i\neq j} \theta_{ij}^{(2)} X_{ij}^{(2)}\\ &+ \sum_{i\neq j} \psi_{ij}^{(1)} X_{ij}^{(1)} X_{ij}^{(2)} + \sum_{i\neq j} \psi_{ij}^{(2)} X_{ij}^{(1)} X_{ji}^{(2)} + \sum_{i\neq j} \psi_{ij}^{(3)} X_{ij}^{(1)} X_{ji}^{(1)} X_{ij}^{(2)}\\ &+ \sum_{i\neq j} \psi_{ij}^{(4)} X_{ij}^{(1)} X_{ij}^{(2)} X_{ji}^{(2)} + \sum_{i<j} \psi_{ij}^{(5)} X_{ij}^{(1)} X_{ji}^{(1)} X_{ij}^{(2)} X_{ji}^{(2)} \Big)\end{aligned} \tag{12}$$

where the parameterization in expression (12) allows for dependence between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ (Holland and Leinhardt, 1981).

Now, analogous to the spirit of expressions (3)–(5), we could imagine all sorts of modelling assumptions on the parameters $\phi$, $\theta$ and $\psi$, including the introduction of covariances between some of the parameters. However, it is advised to exercise caution when modelling as the introduction of too many parameters complicates the interpretation and often leads to situations where the data cannot identify parameters. We take this position in the next section where we are judicious about the introduction of additional parameters.

### 4.1. Corporate law firm example continued

In addition to the advice relationship we now consider the ‘co-working’ relationship whereby a tie is established if a lawyer assisted another lawyer on a project. The co-working relationship is described in more detail in Lazega and Pattison (1999). We denote the advice relationship with the superscript (1) and the co-working relationship with the superscript (2).

We consider a submodel of model (12):

$$\Pr(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \propto \exp\Big( \sum_{i<j} \phi_{ij}^{(1)} X_{ij}^{(1)} X_{ji}^{(1)} + \sum_{i \neq j} \theta_{ij}^{(1)} X_{ij}^{(1)} + \sum_{i<j} \phi_{ij}^{(2)} X_{ij}^{(2)} X_{ji}^{(2)} + \sum_{i \neq j} \theta_{ij}^{(2)} X_{ij}^{(2)} \Big).$$

This implies that the advice and co-working ties are conditionally independent. Dependence is introduced via the assumptions on the $\phi$- and $\theta$-parameters. For $k = 1, 2$, we assume that

$$\begin{aligned} \phi_{ij}^{(k)} &= \phi^{(k)}, \\ \theta_{ij}^{(k)} &= \theta^{(k)} + \alpha_i^{(k)} + \beta_j^{(k)} \end{aligned} \tag{13}$$

but allow that density and reciprocity parameters for the two relationships are correlated. Similarly, the subject-specific random effects are assumed to have an unrestricted variance–covariance structure as follows:

$$\begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \end{pmatrix} \sim \text{normal}\left\{ \begin{pmatrix} \mu_{\theta 1} \\ \mu_{\theta 2} \end{pmatrix}, \Sigma_\theta \right\}, \qquad \begin{pmatrix} \phi^{(1)} \\ \phi^{(2)} \end{pmatrix} \sim \text{normal}\left\{ \begin{pmatrix} \mu_{\phi 1} \\ \mu_{\phi 2} \end{pmatrix}, \Sigma_\phi \right\},$$
$$\begin{pmatrix} \alpha_i^{(1)} \\ \alpha_i^{(2)} \\ \beta_i^{(1)} \\ \beta_i^{(2)} \end{pmatrix} \sim \text{normal}\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_{\alpha\beta} \right\} \tag{14}$$

and $\mu_{\theta 1} \sim \text{normal}(\mu_0, \sigma_0^2)$, $\mu_{\theta 2} \sim \text{normal}(\mu_0, \sigma_0^2)$, $\mu_{\phi 1} \sim \text{normal}(\mu_0, \sigma_0^2)$ and $\mu_{\phi 2} \sim \text{normal}(\mu_0, \sigma_0^2)$ with $\mu_0 = 0$ and $\sigma_0 = 10000$.

The prior distributions on $\Sigma_\theta$, $\Sigma_\phi$ and $\Sigma_{\alpha\beta}$ are inverse Wishart according to standard Bayesian practice and we comment that Wishart distributions are available in WinBUGS. The Wishart distributions are specified by using diagonal matrices with minimal degrees of freedom.

In Table 2, we present some summary results from the multigraph analysis. Our notation is such that $\text{corr}(\theta^{(1)}, \theta^{(2)})$, for example, is the population-specific parameter $\Sigma_\theta(1,2)/\{\Sigma_\theta(1,1) \times \Sigma_\theta(2,2)\}^{1/2}$ where $\Sigma_\theta(i, j)$ is the $(i, j)$th element of the matrix $\Sigma_\theta$. These correlation parameters have their own posterior distributions and the associated entries in Table 2 are the posterior means and standard deviations. We observe that the results for the advice data are comparable with the base model results in Table 1. The estimates from the co-working relationship are such

**Table 2.** Posterior means and standard deviations SD from the multigraph analysis on the corporate law firm data

| *Advice parameters* | | | *Co-working parameters* | | | *Cross-relationships parameters* | | |
|---|---|---|---|---|---|---|---|---|
| *Parameter* | *Mean* | *SD* | *Parameter* | *Mean* | *SD* | *Parameter* | *Mean* | *SD* |
| $\sigma_{\alpha 1}^2$ | 1.08 | 0.22 | $\sigma_{\alpha 2}^2$ | 0.62 | 0.13 | $\text{corr}(\alpha^{(1)}, \alpha^{(2)})$ | 0.49 | 0.10 |
| $\sigma_{\beta 1}^2$ | 0.76 | 0.16 | $\sigma_{\beta 2}^2$ | 1.04 | 0.20 | $\text{corr}(\beta^{(1)}, \beta^{(2)})$ | 0.45 | 0.11 |
| $\text{corr}(\alpha^{(1)}, \beta^{(1)})$ | −0.26 | 0.13 | $\text{corr}(\alpha^{(2)}, \beta^{(2)})$ | −0.59 | 0.10 | $\text{corr}(\alpha^{(1)}, \beta^{(2)})$ | 0.03 | 0.13 |
| $\theta^{(1)}$ | −2.43 | 0.17 | $\theta^{(2)}$ | −2.91 | 0.11 | $\text{corr}(\alpha^{(2)}, \beta^{(1)})$ | −0.24 | 0.14 |
| $\phi^{(1)}$ | 1.80 | 0.13 | $\phi^{(2)}$ | 3.92 | 0.15 | $\text{corr}(\theta^{(1)}, \theta^{(2)})$ | 0.70 | 0.37 |
| EPD | 1223 | 28 | EPD | 1485 | 30 | $\text{corr}(\phi^{(1)}, \phi^{(2)})$ | 0.60 | 0.43 |

that an interpretation can be given to this relationship that is similar to that given to the advice relationship in Section 3. Regarding the correlations that were introduced through the multi-graph approach, we observe that $\mathrm{corr}(\alpha^{(1)}, \alpha^{(2)})$ and $\mathrm{corr}(\beta^{(1)}, \beta^{(2)})$ are moderately positive. It tells us that lawyers who are more likely to seek advice are also more likely to extend co-working ties, and the lawyers who are sought after as co-workers are also popular sources of advice. Almost a null value of $\mathrm{corr}(\alpha^{(1)}, \beta^{(2)})$ means that the tendency to seek advice is not related to attracting co-working ties. A mild negative value for $\mathrm{corr}(\alpha^{(2)}, \beta^{(1)})$ provides some evidence that the ability to attract advice ties is inversely related to the ability to send co-working ties. It seems to fit a natural explanation that those lawyers (perhaps the senior ones) who are sought after more as advisers tend to avoid (or do not encourage) co-working relationships. A high positive value of $\mathrm{corr}(\theta^{(1)}, \theta^{(2)})$ means that among the lawyers the overall tendencies to form ties on the advice and co-working relationships are directly related. Similarly, $\mathrm{corr}(\phi^{(1)}, \phi^{(2)})$ is positive, which tells us the overall tendencies to reciprocate advice and co-working ties move in the same direction.

## Acknowledgements

Both authors were partially supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors thank Emmanuel Lazega of the University of Versailles for providing the law firm data. The authors are grateful to two referees whose comments led to an improvement in the paper.

## References

Anderson, C. J., Wasserman, S. and Crouch, B. (1999) A $p^*$ primer: logit models for social networks. *Socl Netwrks*, **21**, 37–66.
Besag, J. (1975) Statistical analysis of non-lattice data. *Statistician*, **24**, 179–195.
Besag, J. (2000) Markov chain Monte Carlo for statistical inference. *Working Paper 9*. Center for Statistical and the Social Sciences, University of Washington, Seattle.
Daniels, M. J. and Gatsonis, C. (1999) Hierarchical generalized linear models in the analysis of variations in health care utilization. *J. Am. Statist. Ass.*, **94**, 29–42.
Fienberg, S. E. and Wasserman, S. (1981) Categorical data analysis of single sociometric relations. In *Sociological Methodology* (ed. S. Leinhardt), pp. 156–192. San Francisco: Jossey-Bass.
Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995) *Bayesian Data Analysis*. New York: Chapman and Hall.
Gelman, A., Meng, X.-L. and Stern, H. S. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sin.*, **6**, 733–807.
Gill, P. S. and Swartz, T. B. (2001) Statistical analyses for round robin interaction data. *Can. J. Statist.*, **29**, 321–331.
Gill, P. S. and Swartz, T. B. (2003) Bayesian analysis for dyadic designs in psychology. To be published.
Holland, P. W. and Leinhardt, S. (1981) An exponential family of probability distributions for directed graphs. *J. Am. Statist. Ass.*, **76**, 33–65.
Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *J. Am. Statist. Ass.*, **90**, 377–395.
Laud, P. W. and Ibrahim, J. G. (1995) Predictive model selection. *J. R. Statist. Soc.* B, **57**, 247–262.
Lazega, E. and Pattison, P. E. (1999) Multiplexity, generalized exchange and cooperation in organizations: a case study. *Socl Netwrks*, **21**, 67–90.
Nowicki, K. and Snijders, T. A. B. (2001) Estimation and prediction for stochastic blockstructures. *J. Am. Statist. Ass.*, **96**, 1077–1087.
Snijders, T. A. B. (2002) Markov chain Monte Carlo estimation of exponential random graph models. *J. Socl Struct.*, **3**.
Spiegelhalter, D., Thomas, A. and Best, N. (2000) *WinBUGS Version 1.3 User Manual*. Cambridge: Medical Research Council Biostatistics Unit.
Strauss, D. and Ikeda, M. (1990) Pseudolikelihood estimation for social networks. *J. Am. Statist. Ass.*, **85**, 204–212.
Wang, Y. J. and Wong, G. Y. (1987) Stochastic blockmodels for directed graphs. *J. Am. Statist. Ass.*, **82**, 8–19.
Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wasserman, S. and Pattison, P. (1996) Logit models and logistic regression for social networks: I, An introduction to Markov graphs and $p^*$. *Psychometrika*, **61**, 401–425.
Wong, G. Y. (1987) Bayesian models for directed graphs. *J. Am. Statist. Ass.*, **82**, 140–148.