A Bayesian Approach for the Analysis of Triadic Data in Cognitive Social Structures

Tim B. Swartz, Paramjit S. Gill and Saman Muthukumarana *

Abstract

This paper proposes a fully Bayesian approach for the analysis of triadic data in social networks. Inference is based on Markov chain Monte Carlo methods as implemented in the software package `WinBUGS`. We apply the methodology to two datasets to highlight the ease with which cognitive social structures can be analyzed.

Keywords: Bayesian analyses, Cognitive social structures, Markov chain Monte Carlo, Social network analysis, Triadic relations, `WinBUGS`.

*Tim Swartz is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Paramjit Gill is Associate Professor, Department of Mathematics & Statistics, University of British Columbia Okanagan, 3333 University Way, Kelowna BC, Canada V1V1V7. Saman Muthukumarana is Assistant Professor, Department of Statistics, University of Manitoba, 327 Machray Hall, Winnipeg MB, Canada R3T2N2. All authors have been partially supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors thank Phil Rodkin who provided the second dataset concerning schoolyard bullying. The authors also thank the Editor and two anonymous reviewers whose comments led to a substantial improvement in the paper. Finally, the authors thank their colleague Carl Schwarz who helped with a significant programming difficulty.

1 Introduction

The study of social interactions has a rich history in many disciplines including anthropology, sociology, psychology, gender studies and the behavioural sciences. A challenging area related to social interactions concerns interpersonal perceptions, that is, the views that subjects form about the relationships among other subjects, possibly including themselves. The fundamental challenge involves the reconciliation of diverse perceptions.

To motivate the area, consider the friendship relation. We may be interested in the interconnection between an individual's capacity for extending friendship and receiving friendship. For example, does being a friend to someone encourage reciprocal feelings from the other individual? We may also be interested whether subjects perceive their own relationships in the same way as others perceive those relationships. This is an example of a question of "accuracy" which is an obvious yet difficult problem to address (Kenny 1994, Chapter 7). In cognitive social structures, the possibility of bias is real (Krackhardt 1987). A common definition of accuracy is a special type of consensus or interpersonal agreement among a group of judges (Kruglanski 1989). In the absence of a measure of absolute truth or a gold standard, the judgements of others is often used as a criterion.

We study the friendship relation in our first data example. In the second data example, we consider the perception of schoolyard bullying. Of course, there are other cognitive social relations that fall into this class of problems including the seeking of advice, trust in a relationship, the assessment of social status, etc.

Perhaps the most detailed form in which interpersonal perception data arise is referred to as "triadic" data. Triadic data are a complex data structure where responses are collected on a group of subjects. A group of n subjects gives rise to $\binom{n}{2}$ pairs of subjects, and in triadic data, every subject provides two responses on each pair. For example, in the friendship relation, every subject determines whether i considers j a friend, and also whether j considers i a friend, for all $i \neq j$. Note that friendship is a directed relation and that triadic data includes self assessment.

With triadic data, the formulation of realistic and tractable statistical models that address the inferential questions of interest has proven difficult. Many of the suggested approaches to the analysis of triadic data first employ data reduction procedures such as correspondence analysis (Kumbasar 1996) and locally aggregrated procedures (Krackhardt 1987). Consequently, triadic data have not been studied in their full glory, taking into account the complexity and richness of the data structure. The purpose of this communication is to present a computationally tractable, fully Bayesian approach to the analysis of triadic data. We motivate our approach by extending a random effects model where analysis has remained elusive from a classical frequentist point of view. The classical approach has been based on analysis of variance (ANOVA) procedures, an approach known to be cumbersome and unsatisfactory (Lüdtke et al. 2013).

At a superficial level, triadic data seem to share features with paired comparisons data (David 1988) and hence one might consider adapting paired comparisons methodology to the present setting. With triadic data however, directed measurements are taken on a pair of subjects leading to two response variables. In paired comparisons data, there is typically a single measurement on a pair of subjects. Sometimes, paired comparisons data has two responses (e.g. goals for each team in sporting events) from which a difference is taken resulting in a single response. As suggested above, and will be seen later, the inferential questions arising in triadic data settings are typically complex and relate to the perceptions involving judges, senders and receivers. In paired comparisons scenarios, the questions tend to be simpler, and include problems such as the identification of the most preferred subject, the determination of whether subject i is preferred to subject j and the evaluation of the relative ratings given by judges.

Over the past 25 years or so, a number of methodological approaches have been proposed to study the accuracy and consensus of interpersonal perceptions using triadic data structures. Bond, Horn & Kenny (1997) extended the social relations model of Kenny (1994) to incorporate third-person perceptions. Bond et al. (2000) further extended the model to analyze multivariate triadic relations. Their model incorporated covariances between measurements on different types of relations on the same pair of subjects. For example, the model can be used to investigate whether a friendship tie (the first type of relation) has an impact on the tendency to seek advice (the second type of relation). An alternative approach was considered by Koskinen (2004) who extended the model of Batchelder et al. (1997) to the Bayesian setting using prior distributions to overcome nonidentifiability issues. Koskinen (2004) was primarily concerned with the problem of model selection. On the computational side, Butts (2010) and Butts, Hunter & Handcock (2011) have developed a range of social network analysis tools in the programming language R. However, these tools do not include software for the triadic relations models described in this paper.

In section 2, we propose a Bayesian approach for the analysis of binary data in a triadic setting. Various modelling assumptions and the introduction of prior distributions leads to complex hierarchical models. We discuss practical modelling issues including parameter interpretation and prior specification. Some innovative modelling is proposed including the treatment of a covariance matrix with nonstandard form. Following the model for binary data, we outline an analogous model for continuous data. In section 3, computational issues including model selection are discussed. Markov chain Monte Carlo (MCMC) methods are implemented via the software package `WinBUGS` (Spiegelhalter, Thomas, Best & Lunn 2003). `WinBUGS` is freely available from `www.mrc-bsu.cam.ac.uk/bugs`. In sections 4 and 5, Bayesian implementations are illustrated for the two datasets. Some concluding remarks are provided in section 6.

2 A Hierarchical Model

Triadic data structures are extensions of round robin data structures. Round robin data arise when pairs of individuals interact with one another, producing two outcomes for each pair (Gill & Swartz, 2001). With n subjects in a group, a round robin structure therefore has $n(n-1)$ responses. A triadic data structure has an additional peripheral level of assessment where each subject in the group provides $n(n-1)$ responses corresponding to their perceptions of the round robin outcomes. Thus, each of the n subjects serves as a judge to provide their perception of each dyadic relation in the network. For the dyad (pair) of subjects $i \neq j$, let the measurement y_{ij}^k represent the binary response of subject k as a judge about the presence/absence of a directed relation from subject i as an actor towards subject j as a partner (target). More specifically,

$$y_{ij}^k = \begin{cases} 1 \text{ if judge } k \text{ says that a tie exists from subject } i \text{ to subject } j \\ 0 \text{ if judge } k \text{ says that a tie does not exist from subject } i \text{ to subject } j \ . \end{cases}$$

The collection of observations $\{y_{ij}^k\}$ is the social network where $k = 1, \ldots, n$, $i \neq j$. With continuous data, this formulation was referred to as a triadic relations model by Bond, Horn & Kenny (1997).

Using statistical notation, conditional on the p_{ij}^k, we assume independent binary responses $y_{ij}^k \sim \text{Bernoulli}(p_{ij}^k)$. That is, $\text{Prob}(y_{ij}^k = 1) = p_{ij}^k$. The mean structure concerning the presence of a tie from subject i to j as perceived by judge k is given by

$$\begin{aligned} H(p_{ij}^k) &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \\ H(p_{ji}^k) &= \mu + \alpha_j + \beta_i + \gamma_k + (\alpha\beta)_{ji} + (\alpha\gamma)_{jk} + (\beta\gamma)_{ik} \end{aligned} \quad (1)$$

where H is known as a link function and is typically the inverse of a common distribution function. For classical generalized linear models, the logistic link and the probit link are popular choices.

In model (1), μ represents the overall density of ties in the network. The parameters α_i and β_i represent, respectively, the properties of expansiveness (the ability to send ties) and attractiveness (the ability to attract ties) corresponding to actor i. The parameter γ_k is the effect of subject k as a judge which represents his or her "perception bias" to sense the presence of ties in the network. The interaction effect $(\alpha\beta)_{ij}$ represents the special adjustment that subject i has for sending a tie to subject j over and above the expansiveness of subject i. The parameter $(\alpha\gamma)_{ik}$ represents the perception bias of judge k regarding the ability of subject i to send ties. Similarly, the parameter $(\beta\gamma)_{jk}$ represents the perception bias of judge k regarding the ability of subject j to attract ties. This formulation allows us to model the heterogeneity induced by disparate social abilities and by the variability in the judging abilities of the members in the network. It is conceivable that some components of this heterogeneity can be (or should be) explained by some subject-specific factors such as age, sex, social status etc. The model can be easily expanded to accommodate those factors as covariates. Note that model (1) contains non-identifiabilities which may pose difficulties involving interpretation, computation and inference. We have therefore imposed sum-to-zero constraints $\sum \alpha_i = \sum \beta_i = \sum \gamma_i = 0$ on the main effect parameters. However, sum-to-zero constraints corresponding to the interaction effects $(\alpha\beta)_{ij}$ and $(\alpha\gamma)_{ij}$ proved more difficult to implement in `WinBUGS`. We have assigned prior distributions on the interaction effects that are centred about zero and assign appreciable probability only to plausible values of the parameters. This is a convenient tactic that has been used to combat nonidentifiability (Exercise 5.1.e of Leonard & Hsu 1999).

At this stage, we propose a modification to model (1) which improves interpretation and later provides a simplification in the modelling of covariances. The assumption is particularly appropriate when the directional relation is friendship. The assumption is that judge k, in assessing an individual i, will view the individual's expansiveness and attractiveness for ties in the same way. In other words, the judge will tend to be either optimistic or pessimistic concerning the individual forming ties in either direction. In symbols, we assume

$$(\beta\gamma)_{ik} = (\alpha\gamma)_{ik} \ . \tag{2}$$

The consequence of assumption (2) is that the mean structure in (1) becomes

$$\begin{aligned} H(p_{ij}^k) &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\gamma)_{jk} \\ H(p_{ji}^k) &= \mu + \alpha_j + \beta_i + \gamma_k + (\alpha\beta)_{ji} + (\alpha\gamma)_{jk} + (\alpha\gamma)_{ik} \ . \end{aligned} \tag{3}$$

Of particular interest are the cases when $k = i$ and when $k = j$. That is, judge k provides assessments on his or her own relationships in the network. For clarification, we rewrite (3) for the self assessment where $k = i$ as

$$\begin{aligned} H(p_{ij}^i) &= \mu + \alpha_i + \beta_j + \gamma_i + (\alpha\beta)_{ij} + (\alpha\gamma)_{ii} + (\alpha\gamma)_{ji} \\ H(p_{ji}^i) &= \mu + \alpha_j + \beta_i + \gamma_i + (\alpha\beta)_{ji} + (\alpha\gamma)_{ji} + (\alpha\gamma)_{ii} \end{aligned}$$

where we note that the parameter $(\alpha\gamma)_{ii}$ represents the special adjustment by judge i in assessing his/her own capability to establish ties. Further, we define

$$\delta_i = (\alpha\gamma)_{ii} - \frac{1}{n-1} \sum_{k \neq i} (\alpha\gamma)_{ik} \tag{4}$$

which is the difference between subject i's self assessment and the mean assessment of subject i by others. The quantity (4) is an attempt to parametrize the accuracy of self assessment in establishing ties.

In the random effects setting of Bond, Horn & Kenny (1997), α_i, β_i and γ_i follow a trivariate normal distribution with mean vector zero and covariance matrix

$$\Sigma_1 = \begin{bmatrix} \sigma_\alpha^2 & \rho_1\sigma_\alpha\sigma_\beta & \rho_2\sigma_\alpha\sigma_\gamma \\ \rho_1\sigma_\alpha\sigma_\beta & \sigma_\beta^2 & \rho_3\sigma_\beta\sigma_\gamma \\ \rho_2\sigma_\alpha\sigma_\gamma & \rho_3\sigma_\beta\sigma_\gamma & \sigma_\gamma^2 \end{bmatrix} . \tag{5}$$

The variance components σ_α^2 and σ_β^2 represent the variability of subjects to send and receive ties, respectively. The parameter σ^2_γ measures the variability in the perception of members to sense the strength of ties in the network. The correlation parameters in (5) require further description:

- $\rho_1 = corr(\alpha_i, \beta_i)$ measures individual level reciprocity (also called generalized reciprocity) of sending and receiving ties. A positive value of ρ_1 indicates that expansive subjects are those who are relatively more capable of attracting relational ties. A negative value of ρ_1 indicates the presence of asymmetric ties. For example, in networks having hierarchical structures, junior members typically ask seniors for advice and not vice-versa.
- $\rho_2 = corr(\alpha_i, \gamma_i)$ measures one's outlook versus their actual behaviour. A positive value indicates that people who are liberal in establishing ties also perceive that others are liberal in establishing ties. Conversely, people who are conservative in establishing ties perceive that others are also conservative.
- $\rho_3 = corr(\beta_i, \gamma_i)$ also measures one's outlook versus their actual behaviour. A positive value indicates that people who are inclined to attract ties also perceive that others are more likely to attract ties. Conversely, people who are less inclined to attract ties also perceive that others are less likely to attract ties.

For the self-perception bias parameters $(\alpha\gamma)_{ii}$, the random effects model assumes

$$(\alpha\gamma)_{ii} \stackrel{\text{iid}}{\sim} \text{Normal}(0, \sigma^2_{\alpha\gamma}) \ .$$

For the interaction effects in (3) involving reciprocity, it is further assumed in the random effects context that $(\alpha\beta)_{ij}$, $(\alpha\beta)_{ji}$, $(\alpha\gamma)_{ij}$ and $(\alpha\gamma)_{ji}$ follow a multivariate normal distribution with mean vector zero and covariance matrix

$$\Sigma_2 = \begin{bmatrix} \sigma^2_{\alpha\beta} & \phi_1\sigma^2_{\alpha\beta} & \phi_2\sigma_{\alpha\beta}\sigma_{\alpha\gamma} & \phi_3\sigma_{\alpha\beta}\sigma_{\alpha\gamma} \\ \phi_1\sigma^2_{\alpha\beta} & \sigma^2_{\alpha\beta} & \phi_3\sigma_{\alpha\beta}\sigma_{\alpha\gamma} & \phi_2\sigma_{\alpha\beta}\sigma_{\alpha\gamma} \\ \phi_2\sigma_{\alpha\beta}\sigma_{\alpha\gamma} & \phi_3\sigma_{\alpha\beta}\sigma_{\alpha\gamma} & \sigma^2_{\alpha\gamma} & \phi_4\sigma^2_{\alpha\gamma} \\ \phi_3\sigma_{\alpha\beta}\sigma_{\alpha\gamma} & \phi_2\sigma_{\alpha\beta}\sigma_{\alpha\gamma} & \phi_4\sigma^2_{\alpha\gamma} & \sigma^2_{\alpha\gamma} \end{bmatrix} . \tag{6}$$

The variance components $\sigma^2_{\alpha\beta}$ and $\sigma^2_{\alpha\gamma}$ describe the variability in the interaction parameters.

The parameters ϕ_1, ϕ_2, ϕ_3 and ϕ_4 are the corresponding correlations. The interpretation of these correlation parameters is specific to the nature of the relation observed in the network.

Note that without assumption (2), there would have been two additional interaction effects $(\beta\gamma)_{ij}$ and $(\beta\gamma)_{ji}$ resulting in a 6 by 6 matrix Σ_2. This would have led to additional parameters $\phi_5, \ldots, \phi_9$ and $\sigma^2_{\beta\gamma}$. In addition, we would need to jointly model $(\alpha\gamma)_{ii}$ and $(\beta\gamma)_{ii}$.

At this stage, it is worthwhile to pause and take stock of the model. With n subjects in the study, there are $n(n-1)$ relationships, and each relationship is assessed by n judges. This leads to $n^2(n-1)$ observations in the social network. From (3), there are $3n+1$ main effects, $n(n-1)$ parameters corresponding to the interaction terms $(\alpha\beta)_{ij}$ and n^2 parameters corresponding to the interaction terms $(\alpha\gamma)_{ik}$. Additionally, variance component parameters (which are typically the primary parameters of interest in social networks models) are given in (5) and (6). From (5), there are an additional 6 parameters corresponding to Σ_1. From (6), there are an additional 6 parameters corresponding to Σ_2. This leads to a total of $2n(n+1)+13$ model parameters. Therefore for $n \geq 5$, provided that no observations are missing, there are more observations than parameters.

We also wish to emphasize an aspect of the independence assumptions used in modelling the data. Recall that the y^k_{ij} are assumed independent, conditional on the parameters p^k_{ij} which are expressed as functions of other parameters (equation (1)). However, the parameters within (1) are modelled with various multivariate distributions to take into account the dyadic structure. Therefore, there are dependencies in the model which induce marginal dependence amongst all y^k_{ij}. Although complex and high-dimensional, the model as currently proposed can be characterized as a classical random effects model.

2.1 Prior Distributions

We now extend the random effects model by introducing prior distributions that express our uncertainty in the fixed effects parameter μ and the variance component parameters. For the overall density of ties in (3), we assume

$$\mu \sim \text{Normal}(0, 1) \; .$$

Following conventional Bayesian protocol for linear models, we let

$$\Sigma_1^{-1} \sim \text{Wishart}_3[R^{-1}, \nu_0]$$

where $\nu_0 = 3$, $R = I/\nu_0$, I is the identity matrix and $\text{E}(\Sigma_1^{-1}) = \nu_0 R = I$ according to the WinBUGS parametrization of the Wishart. Note that the degrees of freedom parameter $\nu_0 \geq 3$ (the dimension of Σ_1^{-1}) such that Σ_1^{-1} is invertible with probability 1.

The prior distribution for Σ_2 in (6) is problematic due to its unusual structure. For example, it cannot be assumed to have a Wishart form. To specify a prior for Σ_2, we first assign the conjugate priors $1/\sigma^2_{\alpha\beta} \sim \text{Gamma}(100, 10)$ and $1/\sigma^2_{\alpha\gamma} \sim \text{Gamma}(100, 10)$. We then define the matrix $V = \text{diag}(\sigma_{\alpha\beta}, \sigma_{\alpha\beta}, \sigma_{\alpha\gamma}, \sigma_{\alpha\gamma})$ and note that $V'\Phi V = \Sigma_2$ where the correlation matrix Φ is given by

$$\Phi = \begin{bmatrix} 1 & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & 1 & \phi_3 & \phi_2 \\ \phi_2 & \phi_3 & 1 & \phi_4 \\ \phi_3 & \phi_2 & \phi_4 & 1 \end{bmatrix} . \tag{7}$$

The matrix Φ has a nonstandard form where the component parameters ϕ_i have a particular structure. To ensure that Φ is positive definite, we require that the determinants of the three principal minors of Φ are positive. This leads to the following determinant constraints:

$$1 - \phi_1^2 > 0 \tag{8}$$

$$1 - \phi_1^2 - \phi_2^2 - \phi_3^2 + 2\phi_1\phi_2\phi_3 > 0 \tag{9}$$

$$\begin{aligned} & 1 + \phi_2^4 + \phi_3^4 - 2\phi_1\phi_2^2\phi_4 - 2\phi_1\phi_3^2\phi_4 - 2\phi_2^2\phi_3^2 + \phi_1^2\phi_4^2 \\ + \quad & 4\phi_1\phi_2\phi_3 + 4\phi_2\phi_3\phi_4 - \phi_1^2 - 2\phi_2^2 - 2\phi_3^2 - \phi_4^2 \ > \ 0 . \end{aligned} \tag{10}$$

We therefore require a prior distribution on Φ which is defined on the space constrained by (8), (9) and (10). Our proposed prior distribution is simply defined as uniform on the constrained space. In sampling terms, this means that ϕ_1, ϕ_2, ϕ_3 and ϕ_4 are generated independently from the

Uniform$(-1, 1)$ distribution, and only accepted as variates forming the matrix Φ if constraints (8), (9) and (10) are satisfied.

The prior distribution for Φ was introduced primarily for computational and inferential convenience. However, it is reasonable to ask what sort of distribution is this. In Table 1, we provide some summaries concerning the component parameters of Φ which were obtained via simulation. The correlation parameters ϕ_i have prior means of zero and span the interval $(-1, 1)$ with appreciable prior probability. To two decimal places, we also observe that all pairwise correlations involving ϕ_1, ϕ_2, ϕ_3 and ϕ_4 are zero. Therefore, the distribution for Φ appears diffuse, and it may therefore serve as an appropriate prior in the absence of strong prior opinion. An appealing feature of the prior specification is its flexibility. For example, we can easily fix correlation parameters ϕ_i to hypothesized values or consider the equality of various correlation parameters.

Parameter	Prior Mean	Prior Std Dev
ϕ_1	0.00	0.50
ϕ_2	0.00	0.41
ϕ_3	0.00	0.41
ϕ_4	0.00	0.50

Table 1: Prior means and standard deviations for the components of the matrix Φ.

We conclude this subsection with some brief comments about the specified hyperparameters used in the prior distributions of μ, Σ^{-1} and the variance components $\sigma_{\alpha\beta}$ and $\sigma_{\alpha\gamma}$. The first point is that all of these parameters lie fairly deep in the hierarchy, and therefore detailed subjective knowledge concerning these parameters is limited. To give some insight into our hyperparameter choices, we can examine their implications for the probability p of a tie: in the absence of any data for example, the implied prior for p is centred on 0.50 with a 60% probability range from 0.03 to 0.97. This prior therefore allows flexbility in the range of probabilities that can be accommodated.

2.2 Continuous Responses

Now rather than binary random variables y_{ij}^k denoting the presence/absence of ties, there are some network models where continuous responses are observed. For example, judges may be asked to assess the strengths of relationships on a scale of 0 to 100. In this case, we provide the following straightforward modification to the random effects model (3):

$$\begin{aligned} y_{ij}^k &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\gamma)_{jk} + \epsilon_{ijk} \\ y_{ji}^k &= \mu + \alpha_j + \beta_i + \gamma_k + (\alpha\beta)_{ji} + (\alpha\gamma)_{jk} + (\alpha\gamma)_{ik} + \epsilon_{jik} \end{aligned}$$

where the error terms ϵ_{ijk} are assumed independent and normally distributed with mean zero and variance σ_ϵ^2. We assign the standard conjugate prior $1/\sigma_\epsilon^2 \sim \text{Gamma}(1, 1)$. The remainder of the prior specification is the same as in the binary case.

3 Computations

The subjects used in the study are assumed to be a random sample from a population and we are interested in generalizing beyond the particular individuals involved in the study. Putting the modelling components together in their hierarchical structure, the posterior density in the binary response model is therefore proportional to

$$\prod_{i \neq j, k} [y_{ij}^k \mid p_{ij}^k] \; [\mu] \; [\alpha_i, \beta_i, \gamma_i \mid \Sigma_1] \; [\Sigma_1] \; [(\alpha\beta)_{ij}, (\alpha\beta)_{ji}, (\alpha\gamma)_{ij}, (\alpha\gamma)_{ji} \mid \Sigma_2] \; [(\alpha\gamma)_{ii}] \; [\sigma_{\alpha\beta}] \; [\sigma_{\alpha\gamma}] \; [\Phi] \quad (11)$$

where the notation $[\cdot]$ is used to denote a generic probability density function.

Although the posterior distribution given by (11) provides the full description of parameter uncertainty given observed data, the complexity is such that posterior summaries are needed for interpretation. Posterior summaries are typically simple quantities such as posterior means and standard deviations. In this application, these quantities take the form of intractable integrals.

3.1 MCMC Methods

Referring to the posterior density in (11), it seems that a sampling based approach is the only feasible way to approximate posterior summaries. In a Markov chain approach, it is typical to first consider the construction of a Gibbs sampling algorithm. Writing one's own code has some appeal as it facilitates debugging. However, the development of full conditional distributions is a daunting task especially in the case of the logistic link function. Conversely, the writing of `WinBUGS` code is not particularly onerous and allows the user to modify model assumptions. For example, the modification of `WinBUGS` code from the binary response case to the continuous response case involves changing only a few lines of instructions. We remark that there has been some success in handling various types of logit models in Bayesian settings. In Holmes & Held (2006) and Fruhwirth-Schattner & Fruhwirth (2007, 2010), the introduction of auxiliary variables has been used to overcome the difficulty of variate generation from nonstandard full conditional distributions. Along similar lines, Albert and Chib (1993) proposed Gibbs sampling algorithms using data augmentation in the case of probit regression.

We encountered a numerical problem related to the Uniform$(-1, 1)$ distribution used in the specification of the prior for ϕ_1, ϕ_2, ϕ_3 and ϕ_4. With any of the ϕ_i close to -1 or 1, the determinant of the matrix Σ_2 approaches zero. Therefore numerical issues may arise in the calculation of the inverse Σ_2^{-1}, and this causes `WinBUGS` to halt. An easy fix to the problem is to replace the Uniform$(-1, 1)$ distribution with the Uniform$(-0.99, 0.99)$ distribution. There is no logical problem in doing so as we do not expect the absolute values of the correlations to exceed 0.99.

In `WinBUGS`, there sometimes arise programming issues that cause difficulty and whose solution is not well known. The handling of the constraints (9) and (10) proved to be one of these programming difficulties. Whereas interval censoring in `WinBUGS` is typically handled using the I(lower,upper) command, the nature of the constraints in (9) and (10) do not lend themselves to this approach. In such situations, a ”zeros trick” can be used (see www.mrc-bsu.cam.ac.uk/bugs/winbugs/examples/funshapes.txt). For example, to enforce the constraint (9), we use the three `WinBUGS` commands shown below. Here Z is a degenerate Bernoulli random variable set to zero. The second command forces the Bernoulli parameter equal to zero. Finally, the third command implements the constraint (9) by forcing the argument of the step function to be negative.

$$
\begin{array}{ll}
\mathrm{Z} & \leftarrow 0 \\
\mathrm{Z} & \sim \mathrm{dbern(c)} \\
\mathrm{c} & \leftarrow \mathrm{step(phi1} * \mathrm{phi1} + \mathrm{phi2} * \mathrm{phi2} + \mathrm{phi3} * \mathrm{phi3} - 2 * \mathrm{phi1} * \mathrm{phi2} * \mathrm{phi3} - 1)
\end{array}
$$

3.2 Model Selection

A popular method for model selection uses the predictive performance criterion proposed by Laud & 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 outcome y_{ij}^k has probability distribution $p_{ij}^k = \text{Prob}(y_{ij}^k = 1)$. In every MCMC loop, a replicate $y_{ij\text{pred}}^k \sim \text{Bernoulli}(p_{ij}^k)$ can be generated conditional on the currently generated values of the model parameters.

Let $\mathbf{Y}_{\text{pred}}$ denote a replicate of the observed data $\mathbf{Y}_{\text{obs}}$. That is, $\mathbf{Y}_{\text{pred}}$ is generated from a predictive distribution. The particular predictive distribution is the one whose covariates match up to the covariates of $\mathbf{Y}_{\text{obs}}$. In this way, $\mathbf{Y}_{\text{pred}}$ is a replicate of $\mathbf{Y}_{\text{obs}}$. The predictive distribution of $\mathbf{Y}_{\text{pred}}$ under model M is

$$f^{(M)}(\mathbf{Y}_{\text{pred}}|\mathbf{Y}_{\text{obs}}) = \int f(\mathbf{Y}_{\text{pred}}|\eta^{(M)})f(\eta^{(M)}|\mathbf{Y}_{\text{obs}})\, d\eta^{(M)} \tag{12}$$

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

$$\text{E}^{(M)}[d(\mathbf{Y}_{\text{pred}}, \mathbf{Y}_{\text{obs}})|\mathbf{Y}_{\text{obs}}] \tag{13}$$

where $d(\mathbf{Y}_{\text{pred}},\ \mathbf{Y}_{\text{obs}})$ is a discrepancy function and the expectation is with respect to the predictive distribution (12). For binary data, a common discrepancy function is $d(\mathbf{Y}_{\text{pred}}, \mathbf{Y}_{\text{obs}}) = ||\mathbf{Y}_{\text{pred}} - \ \mathbf{Y}_{\text{obs}}||$ which is the total number of incorrect predictions corresponding to model M. It is straightforward to estimate (13) as part of MCMC sampling. In each loop of an MCMC run, $\mathbf{Y}_{\text{pred}}$ is generated and $d(\mathbf{Y}_{\text{pred}},\ \mathbf{Y}_{\text{obs}})$ is calculated. The sample mean of the $d(\mathbf{Y}_{\text{pred}},\ \mathbf{Y}_{\text{obs}})$ values is then used to estimate (13).

In the examples that follow, we also compute the model selection diagnostic DIC known as the Deviance Information Criterion (Carlin & Louis 2009). DIC is a counterpart to the Akaike Information Criterion (AIC) and is an immediate byproduct of sampling from the posterior. In `WinBUGS`, DIC is an option on the *Inference* menu. As with AIC, smaller values of DIC suggest preferred models where differences greater than 3-5 are typically regarded as meaningful.

4 Example 1: Who Likes Whom?

We consider a binary social network which was part of a larger study collected by Krackhardt (1987). Krackhardt collected data of perceptions about interactions among members of a management team (supervisors up through the president) in a high-tech entrepreneurial firm. Each of $n = 21$ persons in the management team was asked to fill out a questionnaire. We consider the friendship response where judge k assessed whether i likes j. An affirmative (negative) response corresponds to $y_{ij}^k = 1(0)$. Although common sense dictates that friendship is a symmetric tie, many non-symmetries have been noted in the real world. A review of 1000 sociometric matrices by Davis and Leinhardt (1972) forced them to alter their model to include non-symmetric ties.

In Figure 1, we provide a graphical summary of the $n^2(n-1) = 8820$ responses where the (i,j)th cell is darker when many of the judges perceive that i likes j. Similarly, the (i,j)th cell is lighter when many of the judges perceive that i does not like j. As expected, we observe considerable symmetry in the matrix which suggests that if i likes j, then in most cases, j also likes i. We also observe that there are various types of personalities in the study. For example, subject 11 tends to like lots of people whereas subject 20 does not have such a happy disposition.

We fit the hierarchical model described in section 2 to the dataset where the logistic link is used. In Table 2, we provide posterior summaries of the variance components and correlation parameters. We first note that most of the parameters are "significant" in the sense that the magnitudes of the posterior means exceed twice the values of the respective posterior standard deviations. We observe that the standard deviations (σ_α, σ_β, σ_γ) corresponding to the main effects are generally larger than the standard deviations ($\sigma_{\alpha\beta}$, $\sigma_{\alpha\gamma}$) of the special adjustment (interaction) parameters. Referring to the correlation parameters in (5), we see that ρ_1 is positive indicating that subjects who like others also tend to be liked. Similarly, and as anticipated, ρ_2 and ρ_3 are also positive. Referring to the correlation parameters in (6), ϕ_1 is nearly 1.0 which implies as expected that there is reciprocity in friendships between particular pairs. Also as anticipated, ϕ_2 and ϕ_3 are positive. Finally, ϕ_4 is close to zero as there appears to be no reason why i's perception of j's ability to make friendships should be related to j's perception of i's ability to make friendships.

With respect to the main effect parameters α_i, β_j and γ_k in (3), the largest posterior means for the α's correspond to subjects 11 and 17. The smallest posterior means for the α's correspond to subjects 20 and 10. This roughly agrees with Figure 1 where darker rows indicate that the individuals like other people. Similar agreement exists with respect to the β parameters and the columns of Figure 1. The general agreement between the data and the posterior means of the α_i and the β_j provide confidence in our model. The posterior means of the α_k parameters vary

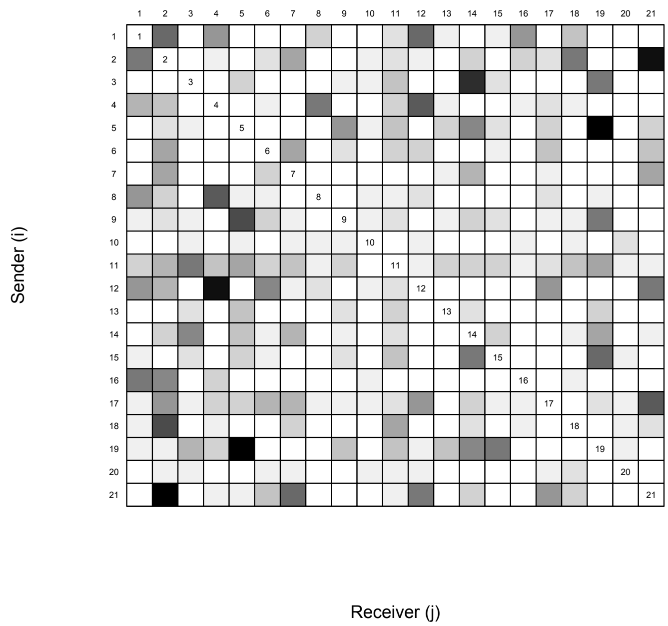

Figure 1: Heatmap illustrating the total scores over all judges with respect to sending ties and receiving ties using the friendship data in Example 1. A darker shade in the (i, j)th cell indicates that many judges perceive that subject i likes subject j. Note that observations are not collected along the diagonal.

from a low of -2.3 to a high of 1.1 indicating a variation in the perception of judges.

We return to the question of accuracy where we ask whether an individual's perception of their relationships is the same as the perception that others hold. In Figure 2, we provide boxplots of the MCMC output corresponding to δ_i in (4) for $i = 1, \ldots, n$. Recall that δ_i represents the adjustment due to the self-perception of individual i with respect to his or her own relationships. In Figure 2, we observe that most subjects have a slightly elevated view of themselves in terms of their capacity to befriend others. Very few individuals have a negative view and there are a few subjects with a grossly inflated view of their ability to form friendship ties.

In an MCMC application, it is good statistical practice to assess practical convergence of the Markov chain. Our summaries in Table 2 are based on 50000 iterations after a burn-in period of 50000 iterations. To gain confidence in our results, we considered parallel runs and

Parameter	Posterior Mean	Posterior Std Dev	95% Credible Interval
σ_α	0.75	0.15	(0.51,1.06)
σ_β	0.91	0.19	(0.62,1.34)
σ_γ	1.19	0.23	(0.83,1.74)
ρ_1	0.53	0.19	(0.10,0.82)
ρ_2	0.25	0.24	(-0.25,0.67)
ρ_3	0.43	0.20	(-0.04,0.77)
$\sigma_{\alpha\beta}$	0.93	0.06	(0.83,1.05)
$\sigma_{\alpha\gamma}$	0.53	0.06	(0.46,0.67)
ϕ_1	0.97	0.02	(0.93,0.99)
ϕ_2	0.41	0.09	(0.21,0.56)
ϕ_3	0.45	0.07	(0.30,0.57)
ϕ_4	0.16	0.17	(-0.12,0.49)

Table 2: Posterior summaries of the variance components and correlation parameters for the friendship data in Example 1.

experimented with different starting values in the Markov chain. Some history and diagnostic plots are provided in the online supplement. For all of the model parameters, the values of the Brooks-Gelman-Rubin statistic $\hat{R}$ (Brooks & Gelman 1998) were less than 1.001 indicating satisfactory convergence.

A Bayesian analysis involving sampling from the posterior provides great flexibility for the analysis of functions of parameters. For example, suppose that we are interested in the difference $(\alpha\beta)_{12} - (\alpha\beta)_{21}$. This represents the reciprocity between sending and receiving ties involving subjects 1 and 2. From posterior sampling, the difference can be calculated for each iteration of the Markov chain. The corresponding posterior density plot is given in Figure 3. As expected with most pairs of subjects, the posterior probability is concentrated about zero. We also observe the range of plausible values of $(\alpha\beta)_{12} - (\alpha\beta)_{21}$.

As part of a larger sensitivity analysis, we considered the effect of the prior distributions $\mu \sim \text{Normal}(0,1)$, $\Sigma_1^{-1} \sim \text{Wishart}_3[R^{-1},3]$, $1/\sigma^2_{\alpha\beta} \sim \text{Gamma}(100,10)$ and $1/\sigma^2_{\alpha\gamma} \sim \text{Gamma}(100,10)$. We modified $\mu \sim \text{Normal}(0,10)$, $\Sigma_1^{-1} \sim \text{Wishart}_3[((\nu_0/5)R)^{-1},5]$, $1/\sigma^2_{\alpha\beta} \sim \text{Gamma}(50,5)$ and $1/\sigma^2_{\alpha\gamma} \sim \text{Gamma}(50,5)$ which has the effect of introducing greater prior dispersion. Under the alternative priors, we observed some changes in the posterior summaries reported in Table 2.

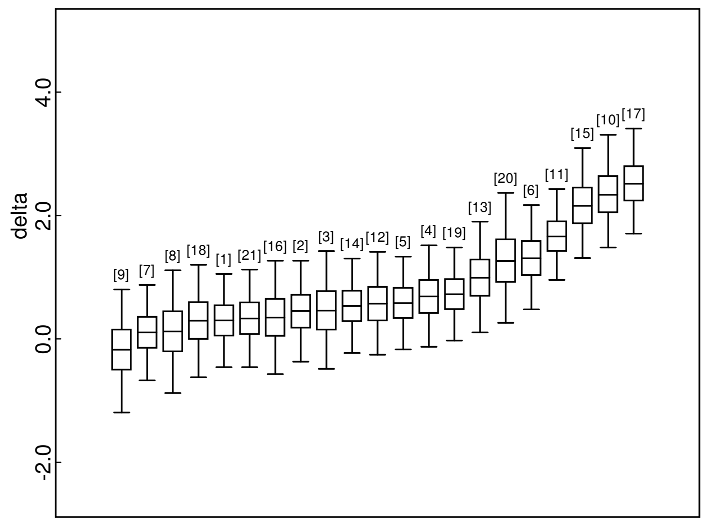

Figure 2: Boxplots of posterior distributions corresponding to the personal assessment parameter δ_i in (4) from Example 1. The label on the top of each boxplot corresponds to the subject number.

For example, the posterior mean for ρ_1 changed from 0.53 to 0.39. The greatest change was observed for the parameter $\sigma_{\alpha\beta}$ whose posterior mean increased from 0.93 to 1.46.

Although we have reported on the full model as described in section 2, it is worth asking whether sub-models are adequate. In Table 3, we record the EPD and DIC diagnostics discussed in section 3.2 for various sub-models. Using either diagnostic, the message is clear that the full model is best at describing the triadic friendship network of Example 1.

5 Example 2: Who Bullies Whom?

Children are routinely classified as bullies and victims but rarely is it known which bullies harass which victims (Rodkin & Berger 2008). The dyadic bully-victim relationship is not easy to observe or measure; a bully may not self identify himself, for example. The subjective experiences of bullying and being bullied may or may not match. Also, bullies and their targets often have different intentions and perceptions of an act (Veenstra et al. 2007). In such a setup,

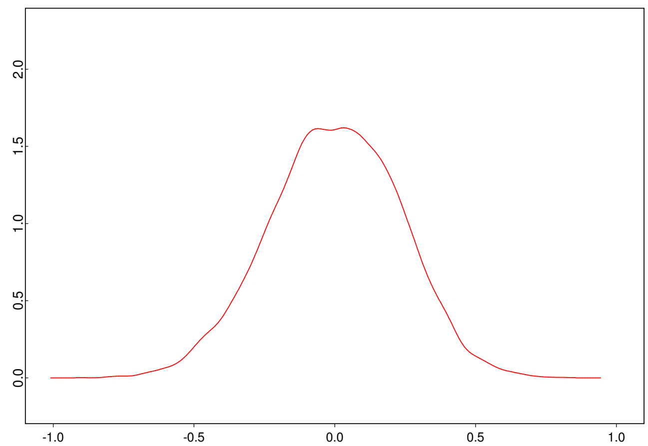

Figure 3: Posterior density plot of $(\alpha\beta)_{12} - (\alpha\beta)_{21}$.

Model	EPD	DIC
μ only	1440.0	5321.2
μ + main effects	1297.0	4631.3
μ + main effects + interaction effects but $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 0$	913.0	3195.0
μ + main effects + interaction effects	920.7	3011.0

Table 3: EPD and DIC diagnostics corresponding to the friendship data in Example 1.

third party observations can be valuable to ascertain and validate the dyadic relationship in the tradition of the social cognitive mapping procedure.

In this example, we do not provide all of the analysis that was performed in the first example. In particular, we do not report on the convergence aspects of the algorithm although we note that no difficulties were encountered. We instead concentrate on the inferential differences between the two examples with an attempt to highlight the flexibility of the proposed methodology.

Card et al. (2010) introduced a triadic relations model in the context of the perception of bullying among school children. Data were collected from third and fourth grade students in public elementary schools in Illinois, USA. A total of 162 children in nine classes participated where each student identified the bullies and their victims from among their classmates. The questionnaire format allowed the respondents to identify themselves as bullies and victims. Very rarely did a subject identify himself or herself as a bully but self-identification as a victim was common.

In this dataset, there was only one possible mechanism for missing data. When students did not participate due to absence from school or lack of parental consent, then they did not fill out the questionnaire. For students who did fill out the questionnaire, there were no missing data as they provided information on all students. In our analysis, we have simply ignored the missing observations, under the assumption that the reasons for missingness were unrelated to the response variable of interest. This enables the missing data to be treated simply as unknown parameters.

The data are binary where

$$y_{ij}^k = \begin{cases} 1 \text{ if student } k \text{ indicates that student } i \text{ bullies student } j \\ 0 \text{ if student } k \text{ does not indicate that student } i \text{ bullies student } j \ . \end{cases}$$

In Figure 4, we depict the bully-victim relationships for a class of 21 children which we subsequently analyze. In this group, 6 children did not provide perception data, and in total, 264 bully-victim ties were reported. We observe that there are relatively few bullies and there seems to be general consensus on the identification of the bullies.

Card et al. (2010) analyzed the full dataset involving 162 children using an ANOVA-based triadic relations model (Bond et al. 1997). With binary responses, we use a more appropriate logit model where we record the bullying behaviour of actor i towards partner j as a potential victim as perceived by judge k

$$\begin{aligned} \text{logit}(p_{ij}^k) &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \\ \text{logit}(p_{ji}^k) &= \mu + \alpha_j + \beta_i + \gamma_k + (\alpha\beta)_{ji} + (\alpha\gamma)_{jk} + (\beta\gamma)_{ik} \ . \end{aligned} \tag{14}$$

In this model, the mean parameter μ measures the overall perception of the bullying behaviour of students in the class. The parameter α_i indicates student i's propensity of bullying others and β_i represents student i's vulnerability of victimization at the hands of others in the class. The parameter γ_i is subject i's "perception bias" to sense the occurrence of bullying in the class.

As in section 2, α_i, β_i and γ_i are assumed to follow a trivariate normal distribution with mean

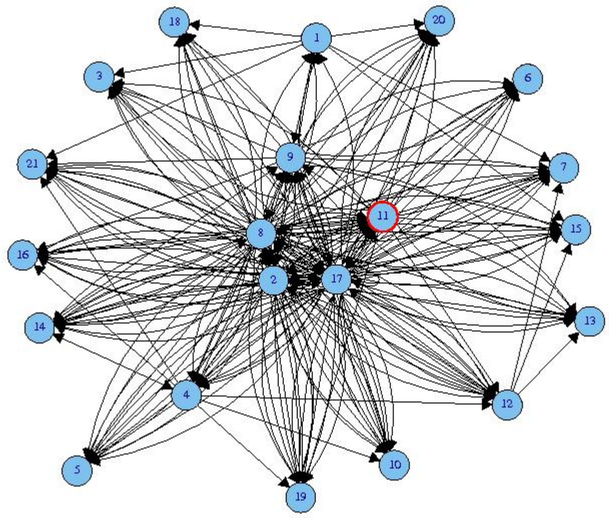

Figure 4: Directed graph showing the reported bully-victim relationships for a class of 21 school children corresponding to Example 2. The number of edges going from student i to student j is the number of judges who reported the relationship. The graph shows the set of bullies in the middle and victims along the periphery of the graph.

vector zero and covariance matrix Σ_1 as specified in equation (5). The variance component σ_α^2 measures the individual variability in aggression - smaller values indicate that all students bully to the same extent whereas larger values indicates that aggressive behaviour can be attributed to a small group of bullies. The component σ_β^2 represents the variability among students as the victims of bullying. Smaller σ_β^2 values indicate that bullies choose their victims from among a large group of fellow students and larger σ_β^2 values indicate that they pick their victims from among a small group of students. The variance parameter σ_γ^2 measures the variability in the perception of students to sense the occurrence of bullying in the class. Smaller values of σ_γ^2 indicate that perception of bullying behaviour is experienced across all the members of the group and larger values mean that perception of bullying is centered only among a small number of students.

Further, we wish to investigate the relationships between the individual traits of the bullying/victimization experience. We formalize these relationships via three correlation parameters whose interpretations are subtle but are important in understanding the behaviours associated with bullying. The three correlation parameters are as follows:

- $\rho_1 = \text{corr}(\alpha_i, \beta_i)$ measures individual level reciprocity of bullying and victimization. A positive value of ρ_1 indicates that students who bully others are likely to be the victims of bullying at the hands of others. A negative value of ρ_1 indicates the presence of asymmetric ties in the sense that bullies are not likely to be bullied and victims usually do not bully others.

- $\rho_2 = \text{corr}(\alpha_i, \gamma_i)$ relates students' perception with their actual bullying behaviour. A positive value indicates that students who bully others are likely to report the bullying behaviour at large (though perhaps not likely to report their own behaviour). A negative correlation means that bullies are not likely to acknowledge the presence of bullying going on in the class.

- $\rho_3 = \text{corr}(\beta_i, \gamma_i)$ measures the relationship between one's outlook versus their victimization experience. A positive value indicates that students who experience bullying also perceive that others are more likely to be victims. Conversely, students who are less likely to attract bullying also perceive that others are less likely to be bullied. A negative correlation is an indication of denial on the part of victims and over-reaction to report bullying on the part of students who are not themselves victimized.

The directed interaction effect $(\alpha\beta)_{ij}$ represents the special propensity of student i bullying student j over and above the bullying potential of subject i and the attraction of subject j to be bullied. The parameter $(\alpha\gamma)_{ik}$ represents the specific perception of judge k regarding the propensity of student i to bully others. On the other hand, the parameter $(\beta\gamma)_{ik}$ represents the specific perception of judge k regarding the vulnerability of student i as a victim of bullying. In contrast to the friendship relation, which is symmetric, we note that the bullying-victimization relation is inherently asymmetric. Therefore, a judge would mostly perceive a subject in opposite directions with respect to bully or victim. In other words, judge k, in assessing an individual i, will view the individual's bullying propensity and victimization vulnerability in the opposite direction. This leads to the assumption

$$(\beta\gamma)_{ik} = -(\alpha\gamma)_{ik} \tag{15}$$

which is the counterpart to the simplification assumption (2). We therefore see that the nature of the relation in triadic data plays a role in the modelling exercise. Under (15), the logit model (14) becomes

$$\begin{aligned}\text{logit}(p_{ij}^k) &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} - (\alpha\gamma)_{jk} \\ \text{logit}(p_{ji}^k) &= \mu + \alpha_j + \beta_i + \gamma_k + (\alpha\beta)_{ji} + (\alpha\gamma)_{jk} - (\alpha\gamma)_{ik} \ .\end{aligned}$$

The study design allows students to identify themselves as possible bullies ($k = i$) and victims ($k = j$). In these cases, some of the parameters have particular importance. For example, the parameter $(\alpha\gamma)_{ii}$ represents subject i's personal perception bias in judging his or her capability to bully others.

We assume that the interaction effects $(\alpha\beta)_{ij}$, $(\alpha\beta)_{ji}$, $(\alpha\gamma)_{ij}$ and $(\alpha\gamma)_{ji}$ follow a multivariate normal distribution with mean vector zero and covariance matrix Σ_2 as in (6). For a detailed interpretation of the variance components $\sigma^2_{\alpha\beta}$ and $\sigma^2_{\alpha\gamma}$ and the correlation parameters ϕ_1, ϕ_2, ϕ_3 and ϕ_4, we refer the reader to Bond et al. (2007) and Card et al. (2010).

Table 4 provides posterior summaries of the variance components and correlation parameters. Among the variance components pertaining to main affects and interaction effects, $\text{var}(\alpha_i) = \sigma^2_\alpha$ has the highest value indicating that most of the variation in the outcomes can be attributed to the heterogeneity in aggressive behaviour due to a small group of bullies. Amongst the correlations ρ_i, only ρ_2, relating students' perception with their actual bullying behaviour, is close to significant. The negative value of ρ_2 is indicative of a bully's tendency to not acknowledge the incidence of bullying and non-bullies' ability to report the bullying behaviour when it is present. To highlight the importance of the nature of the ties in triadic networks, we contrast the highly significant ρ_1 in the friendship data example with the near-zero value of ρ_1 in the bully-victim data. Specifically, friendship is highly reciprocated at the dyadic level but not so with respect to bullying. That is, bullying is not correlated with being bullied.

We carried out a sensitivity analysis for this model changing prior distributions in exactly the same way as was done in Example 1. The value of the posterior mean of ρ_2 changed from -0.40 to -0.54. The largest changes were for the posterior mean of σ_α decreasing from 6.95 to 3.17 and for the posterior mean of $\sigma_{\alpha,\gamma}$ increasing from 0.63 to 1.39.

In a comparison of our results with Table 2 from Card et al. (2010), we note that the largest discrepancy involved the parameter ρ_3. Card et al. (2010) obtained the estimate $\hat{\rho}_3 = -0.91$

Parameter	Posterior Mean	Posterior Std Dev	95% Credible Interval
σ_α	6.95	3.15	(2.58,14.38)
σ_β	0.61	0.13	(0.41,0.92)
σ_γ	1.34	0.38	(0.80,2.27)
ρ_1	-0.02	0.37	(-0.70,0.67)
ρ_2	-0.40	0.36	(-0.90,0.43)
ρ_3	-0.01	0.31	(-0.59,0.59)
$\sigma_{\alpha\beta}$	0.32	0.02	(0.29,0.36)
$\sigma_{\alpha\gamma}$	0.63	0.06	(0.51,0.75)
ϕ_1	-0.21	0.56	(-0.96,0.83)
ϕ_2	0.31	0.19	(-0.04,0.73)
ϕ_3	0.17	0.29	(-0.40,0.68)
ϕ_4	0.73	0.17	(0.31,0.97)

Table 4: Posterior summaries of the variance components and correlation parameters for the bully-victim data in Example 2.

and advised readers that this should be "interpreted cautiously". A strong negative correlation implies that children who are bullied are less likely to perceive bullying. In our analysis, $\rho_3 \approx 0$ which implies a lack of relationship between the experience of being bullied and the perception of bullying. A possible explanation for the discrepancy is that Card et al. (2010) regressed 0/1 response data using ANOVA techniques designed for continuous data.

In Table 5, we present diagnostics for three competing models in the bully-victim data. We observe that the EPD and DIC diagnostics both suggest that the interaction terms are important components in describing the underlying bully-victim mechanism.

Model	EPD	DIC
μ only	503.0	2188.0
μ + main effects	355.7	1347.0
μ + main effects + interaction effects but $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 0$	251.6	874.5
μ + main effects + interaction effects	246.2	876.4

Table 5: EPD and DIC diagnostics corresponding to the bully-victim data in Example 2.

6 Discussion

Assessing accuracy in social relations data has been an elusive problem in the field of social relations analysis. In this paper, we have provided a fully Bayesian analysis for triadic models in both the binary response and the continuous response settings. The models are relatively easy to specify in the `WinBUGS` programming language where the construction of Markov chains is done in the background. A great benefit of the `WinBUGS` implementation is that model modifications may often be accomplished without great difficulty (e.g. drop interaction terms). A user may also choose to modify prior distributions in the light of available prior information. Whereas we have primarily used default priors for the sake of illustration, more informative prior distributions are often beneficial both in terms of convergence and in obtaining stronger inferences. The `WinBUGS` code utilized in this paper may serve as a template for the development of specialized models in the analysis of triadic social relations data.

References

[1] Albert, J.H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 669-679.

[2] Batchelder, W.H., Kumbasar, E. & Boyd, J.P. (1997). Consensus analysis of three-way social network data, *The Journal of Mathematical Sociology*, 22, 29-58.

[3] Bond, Jr., C.F., Horn, E.M., & Kenny, D.A. (1997). A model for triadic relations. *Psychological Methods*, 2, 79-94.

[4] Bond, Jr., C.F., Kenny, D.A., Broome, E.H., Stokes-Zoota, J.J. & Richard, F.D. (2000). Multivariate analysis of triadic relations. *Multivariate Behavioral Research*, 35, 397-426.

[5] Brooks, S.P. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.

[6] Butts, C.T. (2010). Package "sna", http://cran.r-project.org/web/packages/sna/sna.pdf.

[7] Butts, C.T., Hunter, D. & Handcock, M.S. (2010). Package "network", http://cran.r-project.org/web/packages/network/network.pdf.

[8] Card, N.A., Rodkin, P.C. & Garandeau, C.F. (2010). A description and illustration of the triadic relations model: Who perceives whom as bullying whom? *International Journal of Behavioral Development*, 34(4), 374-383.

[9] Carlin, B.P. & Louis, T.A. (2009). *Bayesian Methods for Data Analysis, Third Edition.* Boca Raton: Chapman & Hall.

[10] David, H.A. (1988). *The Method of Paired Comparisons.* New York: Oxford University Press.

[11] Davis, J.A. & Leinhardt, S. (1972). The structure of positive interpersonal relations in small groups. In J. Berger (Ed.), *Sociological Theories in Progress*, Volume 2, 218-251. Boston: Houghton Mifflin.

[12] Fienberg, S.E., Meyer, M.M., & Wasserman, S.S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80, 51-67.

[13] Fruhwirth-Schnatter, S. & Fruhwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis*, 51, 3509-3528.

[14] Fruhwirth-Schnatter, S. & Fruhwirth, R. (2010). Data augmentation and MCMC for binary and multinomial logit models. In: Kneib, T., Tutz, G. (Eds.), *Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir.* Physica-Verlag, Heidelberg, 111-132.

[15] Gill, P.S. & Swartz, T.B. (2001). Statistical analyses for round robin interaction data. *The Canadian Journal of Statistics*, 29, 321-331.

[16] Holmes, C.C. & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1, 145-168.

[17] Kenny, D.A. (1994). *Interpersonal Perception: A Social Relations Analysis.* Guilford Press: New York.

[18] Koskinen, J.H. (2004). Model selection for cognitive social structures. Research Report 2004:3, Department of Statistics, Stockholm University (http://gauss.stat.su.se/site/pdfer/RR2004_3.pdf).

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

[20] Kruglanski, A.W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106, 395-409.

[21] Kumbasar, E. (1996). Methods for analyzing three-way cognitive network data. *Journal of Quantitative Anthropology*, 6, 15-34.

[22] Laud, P.W. & Ibrahim, J.G. (1995). Predictive model selection. *Journal of the Royal Statistical Society B*, 57, 247-262.

[23] Leonard, T. & Hsu, J.S.J. (1999). *Bayesian Methods.* Cambridge University Press: Cambridge.

[24] Lüdtke, O., Robitzsch, A., Trautwein, U. & Kenny, D.A. (2013). A general and flexible approach to estimating the Social Relations Model using Bayesian methods. *Psychological Methods*, 18, 101-119.

[25] Rodkin, P.C. & Berger, C. (2008). Who bullies whom? Social status asymmetries by victim gender. *International Journal of Behavioral Development*, 32, 473-485.

[26] Spiegelhalter, D., Thomas, A., Best, N. & Lunn D.J. (2003). `WinBUGS` *User Manual (Version 1.4.3).* MRC Biostatistics Unit: Cambridge.

[27] Veenstra, R., Lindenberg, S., Zijlstra, B., De Winter, A., Verhulst, F. & Ormel, J. (2007). The dyadic nature of bullying and victimization: testing a dual-perspective theory. *Child Development*, 78, 1843-1854.