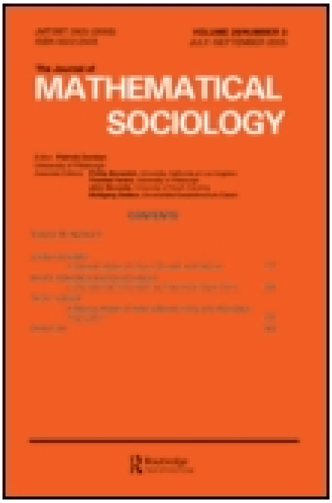

The Journal of Mathematical Sociology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmas20

A BAYESIAN APPROACH TO MODELING STOCHASTIC BLOCKSTRUCTURES WITH COVARIATES

Christian Tallberg [a]
[a] Department of Statistics , Stockholm University
Published online: 03 Sep 2006.

Journal of Mathematical Sociology, 29: 1–23, 2005

ISSN: 0022-250X print/1545-5874 online
DOI: 10.1080/00222500590889703

A BAYESIAN APPROACH TO MODELING STOCHASTIC BLOCKSTRUCTURES WITH COVARIATES

Christian Tallberg
Department of Statistics, Stockholm University

We consider social networks in which the relations between actors are governed by latent classes of actors with similar relational structure, i.e., blockmodeling. In Snijders and Nowicki (1997) and Nowicki and Snijders (2001), a Bayesian approach to blockmodels is presented, where the probability of a relation between two actors depends only on the classes to which the actors belong but is independent of the actors. When actors are a priori partitioned into subsets based on actor attributes such as race, sex and income, the model proposed by Nowicki and Snijders completely ignores this extra piece of information. In this paper, a blockmodel that is a simple extension of their model is proposed specifically for such data. The class affiliation probabilities are modeled conditional on the actor attributes via a multinomial probit model. Posterior distributions of the model parameters, and predictive posterior distributions of the class affiliation probabilities are computed by using a straightforward Gibbs sampling algorithm. Applications are illustrated with analysis on real and simulated data sets.

Keywords: *Bayesian analysis, Blockmodels, Gibbs sampling, Multinomial probit, Random graphs*

1 INTRODUCTION

Most methods in social network analysis are concerned with the description of network structural properties, cf. Wasserman and Faust (1994). One such formal property is *structural equivalence.* A definition is given in Lorrain and White (1971) which, briefly stated, says that two actors are structural equivalent if they have identical relational features. We can then define deterministic approach to blockmodels, first given by White,

Address correspondence to Christian Tallberg, Department of Statistics, Stockholm University, Stockholm. E-mail: christian.tallberg@stat.su.se
The author would like to thank Professor Ove Frank, Dr Mattias Villani, Johan Koskinen and the referees for detailed an insightful comments.

Boorman and Breiger (1976), as a partition of actors into discrete subsets called classes, where actors in the same class are structurally equivalent.

Fienberg and Wasserman (1981) and Holland, Laskey and Leinhardt (1983) generalized the deterministic blockmodel by using the concept *stochastic equivalence*, in a random directed graph model two actors are defined as stochastically equivalent if their probabilistic relation structures to the other actors in the graph are the same. Under an additional assumption of independent dyads and permutation invariance of actors, Fienberg and Wasserman (1981) and Holland, Laskey and Leinhardt (1983) called models with such probabilistic relational structures *stochastic blockmodels*. In the case where the class labels are known, this approach is called a priori block-modeling. When the class labels are unknown, Wasserman and Anderson (1987) proposed a blockmodeling procedure where class labels are identified a posteriori based on the observed relational data within the framework of log-linear models. This specific log-linear model, the p_1 model, introduced by Holland and Leinhardt (1981), includes two parameters for each vertex related to the number of outgoing relations and the number of ingoing relations, as well as the reciprocity parameter. Due to the nature of their model, the range of the parameter space of the two former parameters is limited. Wong (1987) used a Bayesian approach, and computed posterior distributions for the exponential parameters in the p_1 model.

Recent advances in stochastic blockmodels include works by Snijders and Nowicki (1997), Nowicki and Snijders (2001), and Hoff et al. (2002). Hoff et al. (2002) developed a class of models where the probability of a relation between actors depends on the distance in some latent space. Snijders and Nowicki (1997) presented a stochastic a posteriori blockmodel where the number of blocks is restricted to two, and the probability of an edge between two actors depends not only on the classes to which the actors belong but is also independent of the actors. They considered an a posteriori model that is more general than the p_1 model in the sense that the restrictions on the parameter space are not required. In a sequel paper by Nowicki and Snijders (2001), the model is extended to include valued directed graphs where the number of blocks is allowed to be arbitrary.

In many social networks, actors are a priori partitioned into subsets based on actor attributes such as race, sex and income. The model by Nowicki and Snijders (2001) completely ignores such actor information, and the class affiliation probabilities are therefore the same for all actors. Wang and Wong (1987) used a priori information on actor level by proposing a stochastic blockmodel based on p_1, but with actor parameters indicating between-class and within-class tendencies for edges to form.

In this paper, a blockmodel that is an extension of the model by Nowicki and Snijders (2001) is proposed specifically for data where class affiliation depends on actor attributes. The class affiliation probabilities are modeled conditional on the actor attributes via a multinomial probit (MNP) model. Using the simulation-based approach for the MNP model with observable response variables, developed by Albert and Chib (1993), McCulloch and Rossi (1994) and McCulloch et al. (2000), we compute the posterior distributions of the model parameters and predictive posterior distributions of the class affiliation of each actor.

The present paper is structured as follows. In Section 2 the notation is outlined, and the stochastic blockmodel considered is defined. In Section 3, we review the MNP model. Prior distributions and posterior distributions are discussed in Section 4. Section 5 provides empirical examples, and some concluding remarks are given in the final section.

2 NOTATION AND DEFINITION OF THE CONSIDERED STOCHASTIC BLOCKMODEL

Consider a graph of known order v on the vertex set $V = \{1, \ldots, v\}$, and let V^2 denote the set of all distinct ordered pairs of vertices (i, j) from V. We assume a general relational structure on this set of vertices which is represented by its edge value matrix $\mathbf{x} = (x_{ij}), (i,j) \in V^2$, where the element x_{ij} is an observed value of a relation from vertex i to vertex j. By convention, the diagonal entries of $\mathbf{x}$ are equal to 0. Let $x_{ij} \in R$, where $R = \{0, 1, \ldots, r-1\}$ is the range space of the edge values, i.e., the set of possible values of a relation from vertex i to vertex j. In the special cases of graphs and digraphs we have that $R = \{0,1\}$. Furthermore, we assume that V is partitioned into c mutually exclusive non-empty vertex subsets $V_0, V_1, \ldots, V_{c-1}$ called classes, where $|V_k| = v_k, k = 0, \ldots, c-1$ and $v = v_0 + \ldots + v_{c-1}$.

The dyad involving i and j is characterized by $(x_{ij}, x_{ji}, y_i, y_j, \mathbf{z}_i, \mathbf{z}_j)$, where y_i and y_j are class labels of vertices i and j, $\mathbf{z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{ip})'$ and $\mathbf{z}_j = (\mathbf{z}_{j1}, \ldots, \mathbf{z}_{jp})'$ are vectors of known covariate values of vertices i and j in V. Conditional on all $y_1, \ldots, y_v$, the dyads (x_{ij}, x_{ji}) for $i < j$ are independent with probability

$$\Pr(x_{ij} = x, x_{ji} = x' | y_1, \ldots, y_v) = \Pr(x_{ij} = x, x_{ji} = x' | y_i, y_j).$$

If vertex i belongs to class k and vertex j belongs to class l, the dyad probability can be written as

$$\Pr(x_{ij} = x, x_{ji} = x' | y_i = k, y_j = l) = \boldsymbol{\eta}_{kl}(x, x').$$

Define further an array of class dependent dyad probabilities

$$\boldsymbol{\eta} = (\boldsymbol{\eta}_{kl} : 0 \leq k \leq l \leq c-1),$$

where η consists of subarrays

$$\begin{cases} \boldsymbol{\eta}_{kl} = (\boldsymbol{\eta}_{kl}(x,x') : 0 \le x \le r-1, 0 \le x' \le r-1), 0 \le k \le l \le c-1 \\ \boldsymbol{\eta}_{kk} = (\boldsymbol{\eta}_{kk}(x,x') : 0 \le x \le x' \le r-1), 0 \le k \le c-1 \end{cases},$$

satisfying the restriction

$$\begin{cases} \sum_{x=0}^{r-1} \sum_{x'=0}^{r-1} \boldsymbol{\eta}_{kl}(x,x') = 1, 0 \le k < l \le c-1 \\ \sum_{x=0}^{r-1} \sum_{x'=x}^{r-1} \boldsymbol{\eta}_{kk}(x,x') = 1, 0 \le k \le c-1. \end{cases}$$

By symmetry

$$\boldsymbol{\eta}_{kl}(x,x') = \boldsymbol{\eta}_{lk}(x',x) \text{ for all } x, x', k, l,$$

so the model would be over-parametrized by keeping all r^2c^2 dyad probabilities. This is remedied by restricting the parameters to η. Note that there are r^2 elements in η_{kl} for $k < l$ and $\binom{r+1}{2}$ elements in η_{kk} for all k and l. Thus there are $r^2\binom{c}{2} + \binom{r+1}{2}c = \binom{rc+1}{2}$ elements in η and $r^2c^2 - \binom{rc+1}{2} = \binom{rc}{2}$ dyad probabilities are redundant.

To define our stochastic blockmodel, we assume that the class labels y_i are unknown parameters and the prior distribution of the class labels $\mathbf{y} = (y_1, \ldots, y_v)'$ is assumed to depend on the covariate vectors $\mathbf{z}_1, \ldots, \mathbf{z}_v$ according to

$$\Pr(y_1, \ldots, y_v | \mathbf{z}_1, \ldots, \mathbf{z}_v) = \prod_{i=1}^{v} \Pr(y_i | \mathbf{z}_i) = \prod_{i=1}^{v} \theta(y_i | \mathbf{z}_i),$$

where the conditional probability $\Pr(y_i = k | \mathbf{z}_i = \mathbf{z})$ is denoted by $\theta(k|\mathbf{z}_i)$ or by $\theta_k(\mathbf{z}_i)$. Since the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ and η is given by

$$\Pr(\mathbf{x}|\mathbf{y}, \boldsymbol{\eta}) = \prod_{x=0}^{r-1} \prod_{x'=0}^{r-1} \{\eta_{kl}(x,x')\}^{f_{kl}(x,\ x')} \prod_{x=0}^{r-1} \prod_{x'=0}^{r-1} \{\eta_{kk}(x,\ x')\}^{f_{kk}(x,x')},$$

where $f_{kl}(x,x')$ are the counts of dyads $(i,\ j)$ with $i < j$ and $(x_{ij}, x_{ji}, y_i, y_j) = (x, x', k, l)$, the stochastic blockmodel given by the joint distribution of $(\mathbf{x}, \mathbf{y})$, can now be written as

$$\begin{aligned} &\Pr(\mathbf{x}, \mathbf{y} | \boldsymbol{\eta}, \theta\ (y_1|\mathbf{z}_1), \ldots, \theta(y_v|\mathbf{z}_v), \mathbf{z_1}, \ldots, \mathbf{z}_v) \\ &= \left(\prod_{i=1}^{v} \theta(y_i|\mathbf{z}_i) \right) \prod_{x=0}^{r-1} \prod_{x'=0}^{r-1} \{\eta_{kl}(x,x')\}^{f_{kl}(x,x')} \prod_{x=x}^{r-1} \prod_{x'=0}^{r-1} \{\eta_{kk}(x,x')\}^{f_{kk}(x,x')} \qquad (1) \end{aligned}$$

Various properties of the stochastic blockmodel have been studied by, for example, Frank and Harary (1982), Frank (1988a, 1988b) and Janson and Nowicki (1991).

The model given by (1) is a version of the model presented by Nowicki and Snijders (2001), extended to include vertex attributes affecting the class affiliation probabilities. Thus, unlike the simpler model of Nowicki and Snijders (2001), where the probability to belong to class k, θ_k, is the same for all vertices, our proposed model is richer since it allows the class affiliation probabilities, $\theta_k(\mathbf{z}_i)$, to vary between the vertices. When data are generated by (1), we shall show by a simulated example that the prediction of $\mathbf{y}$ is improved considerably. In the sequel, the model by Nowicki and Snijders (2001) is denoted M_1, and the model given by (1) is denoted M_2.

The class affiliation probabilities are modeled conditional on the covariates via the MNP model introduced by Aitchison and Bennet (1970). A brief review of the MNP model is given in the next section.

3 THE MULTINOMIAL PROBIT MODEL

Consider a random utility model in the following way. For the ith individual faced with c choices, suppose that the utility of choice k is

$$w_{ik}^* = \mathbf{z}_i'\beta_k^* + \varepsilon_{ik}^*,\ k = 0, \ldots, c-1,$$

where $\beta_k^* = (\beta_{k0}, \ldots, \beta_{kp})'$ is a $(p+1)$ dimensional parameter vector, and ε_{ik}^* is an unobserved random variable. If the individual makes choice k, we assume that w_{ik}^* is, the maximum among the c utilities. By letting $\mathbf{w}_i^* = \left(w_{i0}^*, \ldots,\ w_{i(c-1)}^*\right)'$ be a vector of c utility indexes, each class label is then a function of $\mathbf{w}_i^*$ as follows

$$y_i(\mathbf{w}_i^*) = \arg \max_{k=0,\ldots,c-1}\ w_{ik}^*.$$

It is conventional to measure utility relative to the alterantive w_{i0}^* to identify the model parameters (McCulloch and Rossi, 1994). Thus, we will normalize by reducing w_i^* to the $(\mathrm{c} - 1)$ dimensional vector

$$\mathbf{w}_i = [w_{ik}^* - w_{i0}^*, k = 1, \ldots, c-1].$$

The utility of choice k relative to choice 0 is then given by

$$w_{ik} = \mathbf{z}_i'\beta_k + \varepsilon_{ik},\ k = 1, \ldots, c-1,$$

where $\beta_k = \beta_k^* - \beta_0^*$, and $\varepsilon_{ik} = \varepsilon_{ik}^* - \varepsilon_{i0}^*$, and each class label is then a function of $\mathbf{w}_i$ given by

$$y_i(\mathbf{w}_i) = \begin{cases} 0 & \text{if} \max_k w_{ik} \leq 0 \\ \arg\max_k w_{ik} & \text{if} \max_k w_{ik} > 0. \end{cases}$$

Thus, $y_i = 0$ if all the w_{ik} are non-positive, otherwise y_i equals the index of the biggest positive w_{ik}. If $\mathbf{w}_i$ is a continuous random vector, the probability that at least two elements are equal is zero so we need not consider ties.

Furthermore, let $\Sigma^* = [\sigma_{kl}, k, l = 0, \ldots, c-1]$ be a c by c positively semi-definite matrix of covariances for the utility of choices. Without loss of generality, we set the first row and column of Σ^* to zeros and refer to the remaining submatrix as

$$\Sigma = [\sigma_{kl}, k, l = 1, \ldots, c-1].$$

Normally, the choice provided by the greatest utility is observable, whereas in our setting the choices, signifying class labels, are unobservable. We also note that Σ or β must be normalized in some way because the scale of the distribution of $\mathbf{w}_i$ is not idenitfiable. A discussion of how to handle this problem is briefly given in Section 4.5.

Assume that $\mathbf{w}_i \sim N(\mathbf{z}_i'(\beta_1, \ldots, \beta_{c-1}), \Sigma)$. If the parameter vectors $\beta_1, \ldots, \beta_{c-1}$ are replaced by a single parameter vector $\beta = (\beta_1', \ldots, \beta_{c-1}')'$, the model can be given as $\mathbf{w}_\mathrm{i} \sim N(\mathbf{Z}_i\beta, \Sigma)$, where

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_i' & & & 0 \\ & \mathbf{z}_i' & & \\ & & \ddots & \\ 0 & & & \mathbf{z}_i' \end{bmatrix}$$

is a matrix of dimension $(c-1)$ by $(c-1)\,p$.

In the special case where the covariates have no effect, the MNP model is reduced to comprise $c - 1 + \binom{c}{2}$ parameters, $c-1$ intercepts and $\binom{c}{2}$ covariances for the utility of choices. Thus, we have $c - 1 + \binom{c}{2}$ effective parameters but $p(c-1) + \binom{c}{2}$ formal parameters. This could lead to difficulties in the estimation procedure. However, extensive simulation results not presented here show that for graphs of large sizes the Gibbs sampler succeed to converge although the formal parameters outnumber the effective parameters in the MNP model irrespective of how informative the specified prior distributions are. For graphs of small sizes though it is necessary to use weakly informative priors on the MNP model parameters to prevent the Gibbs sampler from failing to converge. See Section 5 for further discussion on this subject where our approach is performed on two simulated data sets of which one is generated by a model where all the MNP model parameters except the $\binom{c}{2}$ covariances equal zero.

An alternative to the MNP model would be the multinomial logit model with uncorrelated choices (Koop and Poirier, 1993).

4 PRIOR ASSIGNMENTS AND COMPUTATION OF POSTERIOR DISTRIBUTIONS

The M_2 model includes the set of unknown parameters $\mathbf{y}, \boldsymbol{\eta}, \beta$ and Σ, which require specification of a prior distribution, $p(\mathbf{y}, \boldsymbol{\eta}, \beta, \Sigma)$ according to a Bayesian analysis. The computation of the posterior distribution is given by

$$p(\mathbf{y}, \boldsymbol{\eta}, \beta, \Sigma|\mathbf{x}, \mathbf{z}) \propto p\,(\mathbf{x}, \mathbf{z}|\mathbf{y}, \boldsymbol{\eta}, \beta, \Sigma)\, p\,(\mathbf{y}, \boldsymbol{\eta}, \beta, \Sigma) \tag{3}$$

where $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_v)$ is a p by v matrix of the v covariate vectors. An η is independent of β and Σ conditional on $\mathbf{y}$, and for simplicity independence is assumed between β and Σ, the prior distribution can be decomposed into

$$\begin{aligned} p(\boldsymbol{\eta}, \mathbf{y}, \beta, \Sigma) &= p(\boldsymbol{\eta}|\mathbf{y}, \beta, \Sigma)p(\mathbf{y}|\beta, \Sigma)p(\beta|\Sigma)p(\Sigma) \\ &= p(\boldsymbol{\eta}|\mathbf{y})p(\mathbf{y}|\beta, \Sigma)p(\beta)p(\Sigma) \end{aligned}$$

The posterior joint distribution given by Equation (3) is highly intractable, but since the full conditional posterior distribution of each involved parameter is easy to compute, a feasible approach is to implement the Gibbs sampler algorithm. The Gibbs sampler works by iteratively drawing values from each of the full conditional distributions, each conditionally on the last updated values of all the other unknown parameters. As the number of draws approaches infinity, the Gibbs sampler generates accurate samples from the joint posterior distribution. A more extensive review of the Gibbs sampler algorithm can be found in for example Gelman et al. (1995), and Gilks, Richardson and Spiegelhalter (1996).

4.1 Full Conditional Posterior of y

By following the notation of Nowicki and Snijders (2001), we first define $d_{il}(x, x')$ to be the number of vertices in class l that have relation (x, x') with vertex $i \in V, y_i = k$, which can be expressed more formally as

$$d_{il}(x, x') = \Sigma_{j:(i,j)\in V^2} I\{x_{ij} = x,\ x_{ji} = x'\} I\{y_j = l\}.$$

The full conditional posterior distribution of eacy y_i is then given by

$$\Pr\,(y_i = k|\{y_i\}_{j\neq i}, \eta, \beta, \mathbf{\Sigma}, \mathbf{x}, \mathbf{z}_i) \propto\, \theta_k(\mathbf{z}_i) \prod_{l=0}^{c-1} \{\eta_{kl}(x, x')\}^{d_{il}(x,x')}.$$

4.2 Full Conditional Posterior of the Dyad Probabilities

To determine the full conditional posterior distribution of each set of dyad probabilities, we first define an array of counts of dyad values

$$\mathbf{f} = (\mathrm{f}_{kl} : 0 \le k \le l \le c - 1),$$

where $\mathbf{f}$ consists of

$$\begin{cases} \mathbf{f}_{kl} = (f_{kl}(x, x') : 0 \le x \le r - 1, 0 \le x' \le r - 1), 0 \le k \le l \le c - 1 \\ \mathbf{f}_{kk} = (f_{kk}(x, x') : 0 \le x \le x' \le r - 1), 0 \le k \le c - 1 \end{cases},$$

and an array of hyperparameters

$$\mathbf{a} = (\mathbf{a}_{kl} : 0 \le k \le l \le c - 1),$$

where a consists of

$$\begin{cases} \mathbf{a}_{kl} = (a_{kl}(x, x') : 0 \le x \le r - 1, 0 \le x' \le r - 1), 0 \le k \le l \le c - 1 \\ \mathbf{a}_{kk} = (a_{kk}(x, x') : 0 \le x \le x' \le r - 1), 0 \le k \le c - 1 \end{cases}.$$

By assuming that each set of counts of dyad values is multinomially distributed, and each set of dyad probabilities is conjugate Dirichlet a priori, the full conditional posterior distribution of each set of dyad probabilities is

$$(\boldsymbol{\eta}_{kl} | \mathbf{y}, \beta, \boldsymbol{\Sigma}, \mathbf{x}, \mathbf{Z}) = (\boldsymbol{\eta}_{kl} | \mathbf{f}_{kl}) \sim \text{ Dirichlet } (\mathbf{f}_{kl} + \mathbf{a}_{kl}), k \le l.$$

Note that under vague priors, our model is invariant to permutation of the block labels and is therefore unidentified, see, e.g., Richardson and Green (1997). In the literature, this phenomenon is called label switching, and it causes difficulties to assess accurate posterior distributions. This problem is discussed in Nowicki and Snijders (1997), who suggest that identifiability restrictions are imposed on the elements in $\boldsymbol{\eta}$.

4.3. Full Conditional Posterior of $\mathbf{w}_i$

The main obstacle in implementation of the MNP model has been computing the multivariate normal probabilities for any dimension higher than 2. However, vast improvements of computer-based methods in recent years, such as Gibbs sampling, have made estimation of the MNP model feasible. In this paper, we use the simulation based approach for the MNP, developed by Albert and Chib (1993), McCulloch and Rossi (1994) and McCulloch et al. (2000) in order to compute the posterior distributions of the MNP parameters β and Σ.

First we introduce v independent latent vectors $\mathbf{w}_i$ as discussed in Section 3. The full conditional posterior of each $\mathbf{w}_i$ is equal to the $N(\mathbf{Z}_i\beta, \Sigma)$ distribution, $i = 1, \ldots, v$, truncated to the region

$$\mathbf{w}_i \in \mathcal{R}^{c-1} : \max_k \; w_{ik} \leq 0,$$

if $y_i = 0$, and

$$\mathbf{w}_i \in \mathcal{R}^{c-1} : \arg\max_k w_{ik} = y_i,$$

otherwise. Here $\mathcal{R}$ denotes the set of real numbers. For a detailed description of simulation from a truncated multivariate normal distribution, see Geweke (1991). The draws of $\mathbf{w}_i$ are of no interest per se and need not be saved after the termination of the iteration. They are only introduced to facilitate the computation of the posterior of β and Σ.

4.4. Full Conditional Posterior of β

By selecting a proper conjugate $N(\beta^*, \mathbf{B}^*)$ prior for β, the full conditional posterior of β is $N(\tilde{\beta}, \tilde{\mathbf{B}})$, where

$$\tilde{\beta} = (\mathbf{B}^{*-1} + \mathbf{Z}'(\mathbf{I}_v \otimes \Sigma^{-1})\mathbf{Z})^{-1}(\mathbf{B}^{*-1}\beta^* + \mathbf{Z}'(\mathbf{I}_v \otimes \Sigma^{-1})\mathbf{w}),$$
$$\tilde{\mathbf{B}} = (\mathbf{B}^{*-1} + \mathbf{Z}'(\mathbf{I}_v \otimes \Sigma^{-1})\mathbf{Z})^{-1},$$

$\mathbf{w} = (\mathbf{w}_1', \ldots, \mathbf{w}_v')$ is a $v(c-1)$ dimensional vector, $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_v)$ is a $(p+1)$ by v matrix, $\mathbf{I}_v$ is a v by v identity matrix, and $\mathbf{I}_v \otimes \Sigma^{-1}$ is a $v(c-1)$ by $v(c-1)$ matrix called the Kronecker product of $\mathbf{I}_v$ and Σ^{-1} defined by

$$\begin{bmatrix} \mathbf{I}_{c-1}\Sigma^{-1} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{0} & \cdot & & \\ \cdot & & \cdot & \\ \mathbf{0} & & & \mathbf{I}_c\Sigma^{-1} \end{bmatrix}$$

where $\mathbf{I}_{c-1}$ is a $(c-1)$ by $(c-1)$ identity matrix

4.5 Full Conditional Posterior of Σ

Due to various solutions of the identification problem that arises in the multinomial probit model, there are various possibilities how to set priors on Σ and derive full conditional posteriors. One solution is to condition on $\sigma_{11} = 1$ and construct a Gibbs sampler from the conditional posterior of $\Sigma|\sigma_{11} = 1$ (McCulloch et al. 2000). With this approach it is possible to specify a truly diffuse or improper prior. However, as pointed out by

(McCulloch et al. 2000), this simple method of achieving identification comes at a cost. Due to a Gibbs sampler that produces a Markov chain which tends to be relatively high autocorrelated, the Markov chain will fail to coverage in some extreme cases. These cases occur in high dimensions and in situations in which the likelihood is not very informative. Our preliminary findings for the case of unknown block labels, specifically in graphs of small sizes, are that it is in general difficult for the Gibbs sampler to converge with the identification approach. Instead we conveniently consider a proper conjugate Wishart prior distribution (McCulloch and Rossi, 1994) on $\mathbf{\Sigma}^{-1}$, $\mathbf{\Sigma}^{-1} \sim$ Wishart (m, D), which combined with the likelihood yield the following full conditional posterior distribution

$$\mathbf{\Sigma}^{-1}|\beta, \mathbf{w} \sim \text{Wishart}\big(m + v,\ D + \mathbf{\Sigma}_{i=1}^{v} \varepsilon_i \varepsilon_i'\big),$$

where the Wishart prior distribution is parametrized so that $E(\Sigma^{-1}) = mD$.

Note that as in Nowicki and Snijders (2001), our goal is to compute the predictive posterior distributions of the class labels. In both cases, this is achieved by performing a Gibbs sampling. However, in Nowicki and Snijders (2001), the class affiliation probabilities are equal for all actors, whereas here they depend on actor attributes via the MNP model and are therefore not necessary equal. Hence, we need to estimate an extra set of parameters, β and $\mathbf{\Sigma}$. One way is to adopt the procedure developed by for example McCulloch and Rossi (1994), and McCulloch et al. (2000), who performed a Gibbs sampling for a *fixed* vector of class labels in order to estimate β and $\mathbf{\Sigma}$. In our applications, the vector of class labels is unknown. Thus, in each iteration of the Gibbs sampler, where predictive posterior distributions of the class labels are computed, we have to perform an embedded Gibbs sampling, such as the one suggested by McCulloch and Rossi (1994), and McCulloch et al. (2000), in order to obtain accurate estimated posterior distributions of β and $\mathbf{\Sigma}$ from which we sample the last updated values. However, simulation results show that we obtain the same posterior distributions by performing just one iteration in the embedded Gibbs sampler. A probable explanation is that the Gibbs sampler typically moves rather slowly in y-space.

5 INFERENCE AND MODEL ASSESSMENT ILLUSTRATED WITH NUMERICAL EXAMPLES

We now illustrate the methodology presented in the previous sections using one data set generated by computer simulation and one real data set from the social network modeling literature. The number of classes is predetermined to three, $c = 3$. For simplicity, we will in the given examples consider posterior blockmodeling for undirected graphs. Then the range space of dyad values is reduced to $\{(0,0),(1,1)\}$, and $\eta_{kl}(1,1)$, $f_{kl}(1,1)$ and $a_{kl}(1,1)$ are denoted η_{kl}, f_{kl} and $a_{kl}, 0 \leq k \leq l \leq 2$, respectively. The full conditional posterior distribution of each set of dyad probabilities is then

$$(\boldsymbol{\eta}_{kl}|\mathbf{f_{kl}}), \sim \text{beta}(\mathbf{f}_{kl} + \mathbf{a}_{kl}), 0 \leq k \leq l \leq 2.$$

Each run of the Gibbs sampler concerning the MNP parameters was started at the initial values $\beta = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{I}_2$, whereas the initial values of each element in $\mathbf{y}$ were drawn from a discrete uniform distribution. Since we draw $\mathbf{w}$ first in the part of the Gibbs sampler concerning the MNP parameters, there is no need to specify initial values. For the same reason, there is no need to specify initial values for η.

The information extracted from the Gibbs sampler is only valid for inference as long as the chain has converged. Thus, an important issue is how to detect convergence. Several ad hoc methods have been suggested for determining the chain long enough so that the Gibbs sampler has converged. A general strategy is to monitor the convergence of some aspect of the Gibbs sequence. Gelfand and Smith (1990) and Gelfand, Hills, Racine-Poon and Smith (1990) suggest monitoring density estimates from independent sequences. Tanner (1991) suggests monitoring a sequence of weights that measures the discrepancy between the sampled and the desired distribution. Here, a simple and perhaps naive strategy for assessing convergence is adopted. In each iteration, the mean of the sampled observations is monitored, allowing the sampler to run until we feel that the marginal posterior distributions of the parameters of interest have converged. In all analysis, convergence were considered to have taken place after 5,000 iterations, and computations of posterior distributions are based on samples of additional 10,000 observations.

5.1 Simulated Example

We will use a model including one predictor and intercept. Data are simulated as follows. For $v = 40$, $c = 3$ and $p = 1$, a vector-valued observation of 40 attribute values is generated, where each element is drawn iid from a uniform distribution on the interval $(-2, 2)$. Since we only use one attribute and an intercept, each $\mathbf{z}_i$ is a vector of dimension two. Given the design matrix $\mathbf{Z}_i$, 40 random vectors of utility indexes $\mathbf{w}_i$ is drawn iid from a N ($\mathbf{Z}_i\beta$,Σ)- distribution for two sets of MNP model parameters. In each set, the variances are $\sigma_{11} = 1$ and $\sigma_{22} = 2$, and the corrrelation is $\rho_{12} = 0.5$.

The two sets of β, the corresponding outcome of the randomly generated utility indexes and their associated class labels y_i ($\mathbf{w}_i$) are as follows:

1. β'=(β_{10}, β_{11}, β_{20}, β_{21}) = (-1, -1.4, -0.5, 1.0), where β_{10} and β_{20} are the intercept coefficients and β_{11} and β_{21} are the attribute coefficients. For 12 $\mathbf{w}_i, \max_k w_{ik} \leq 0$ yielding y_i $(\mathbf{w}_i) = 0$, for 14 $\mathbf{w}_i$, $w_{i1} > 0$ is the maximum element yielding $y_i(\mathbf{w}_i) = 1$, and for the remaining 14 $\mathbf{w}_i$, $w_{i2} > 0$ is the maximum element yielding $y_i(\mathbf{w}_i) = 2$. The vector of class labels $\mathbf{y}$ are than rearranged so that the first 12 vertices belong to class 0, the following 14 belong to class 1 and the last 14 belong to class two.
2. $\beta' = (0.0,\ 0.0,\ 0.0,\ 0.0) = 0$. Analogously, the following outcome is obtained: $y_i(\mathbf{w}_\mathrm{i}) = 0$ for 13 vertices, $y_i(\mathbf{w}_i) = 1$ for 11 vertices, and $y_i(\mathbf{w}_i) = 2$ for 16 vertices.

Given $\mathbf{y}$, an adjacency matrix of order 40 is then generated where the off-diagonal elements are drawn independently from Bernoulli distributions with edge probabilities

$$\begin{aligned}\boldsymbol{\eta} &= \{\eta_{00}, \eta_{01},\ \eta_{02},\ \eta_{11},\ \eta_{12},\ \eta_{22}\}\\ &= \{0.6, 0.1, 0.1, 0.4, 0.1, 0.2\}.\end{aligned}$$

Since we only consider proper prior distributions for the MNP model parameters, the situation in which the investigator has no strong prior beliefs about the location of the parameters may be approximated by specifying extremely diffuse priors. This is achieved by taking the precision matrix of β to be very small and by setting m to a small number relative to sample size used in the analysis, which ensures that the prior on $\mathbf{\Sigma}^{-1}$ remains diffuse relative to the likelihood. Extensive simulation results show that our implemented simulation based estimation method yield computed posterior distributions of high accuracy by using extremely diffuse priors for graphs of sizes $v > 100$. However, since realistic sizes of graphs that arise in studies of networks are usually smaller, our reported analysis is performed on a graph of size $v = 40$. To keep the Gibbs sampler from wandering too far a field in the parameter space, we are forced to use slightly informative prior distributions on β and $\mathbf{\Sigma}^{-1}$. Hence, we assume a priori that β is N ($\beta *' = 0$, $\mathbf{B}* = 100\mathbf{I}_4$), and $\mathbf{\Sigma}^{-1}$ is Wishart ($m = 10$, $D = \mathbf{I}_2$) which is centered on the identity matrix. An extensive discussion of the choice of priors on MNP model parameters, and their impact on the inference procedure are given in McCulloch and Rossi (2000). In some situations, briefly mentioned in Section 4.5, McCulloch and Rossi also advocate weakly informative priors on $\mathbf{\Sigma}^{-1}$, centered on the identity matrix, in the absence of strong prior information, since the Gibbs sampler may experience convergence problem.

Marginal posterior distributions of the MNP model parameters are displayed in Figure 1. Not surprisingly, the marginal posterior distribution of Σ^{-1} obtained from the Gibbs sampler is centered on $\mathbf{I}_2$ since our rather informative prior on Σ^{-1} is centered on $\mathbf{I}_2$. Hence, the true value of ρ_{12} just falls outside the 95% credibility interval. It is further striking in Figure 1 how robust our approach to blockmodels with covariates seems to be to the problem of performing inference in an overparametrized model. The credibility intervals are conclusive except for the intercepts.

To interpret the impact of the MNP model parameters on the probabilities $\Pr(y_i = k), k = 0, 1, 2$ as $\mathbf{z}_i$ changes, recall that the latent random utility vector $\mathbf{w}_i$ is $N(\mathbf{Z}_i'\beta, \Sigma)$, and that the class labels are functions of $\mathbf{w}_i$ according to Equation (2) . The preference to the kth class of the ith actor is determined by $\mathbf{z}_i$ and the parameters β_{11}, β_{21} and Σ that depend on the class. For example, the actor age may be regarded as a covariate affecting class affiliation. The corresponding parameters $\beta_{11} = -1.4, \beta_{21} = 1.0$ and

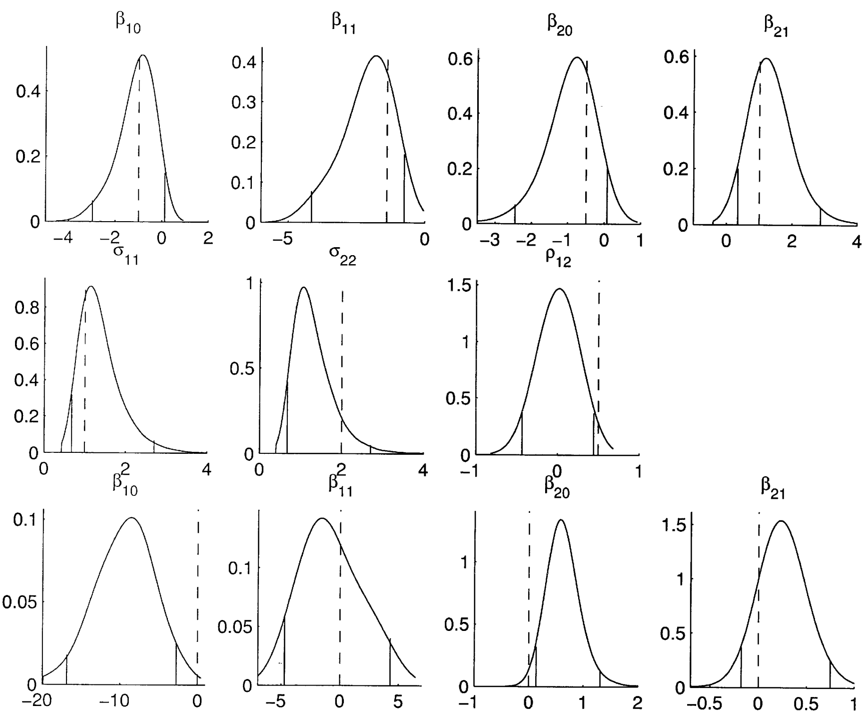

FIGURE 1 Marginal posterior distributions of the MNP model parameters for $\beta' = (-1.0, -1.4, -0.5,\ 1.0)$(the upper two rows), and for $\beta' = \mathbf{0}$ (the bottom row). A dashed line is drawn at the true parameter value, and solid vertical lines depict the lower and upper boundaries of 95% credibility intervals.

$\rho_{12} = 0.5$ depend on the class, e.g. given that ρ_{12} is positive (although we a priori force it to take values near 0) and due to the negativity of β_{11} and the positivity of β_{21}, for increasing age, an increasing probability to belong to class 2 in relation to class 0 but a decreasing probability to belong to class 1 in relation to class 0 may be expected. If instead ρ_{12} should be negative, large values of w_{i1} yield small values of w_{i2}, and vice versa. Then, due to the negativity of β_{11} and the positivity of β_{21}, for increasing age, an increasing probability to belong to class 1 in relation to class 0 should be expected.

In our analysis we avoid making choices about $\boldsymbol{\eta}$ by setting the hyperparameters in the beta distributions to one. That is, all the elements in $\boldsymbol{\eta}$ are uniformly distributed. To avoid the problem of nonidentified class labels, the domain of $\boldsymbol{\eta}$ is restricted to $0 < \eta_{22} < \eta_{11} < \eta_{00} < 1$. Hence, η_{00}, η_{11} and η_{22} are dependent, and defined in the region $0 < \eta_{22} < \eta_{11} < \eta_{00} < 1$ independently of η_{01}, η_{02} and η_{12}. However, if prior information on the elements in η is available, besides yielding improved estimators, a non-uniform prior distribution can identify the class labels and thereby avoid the problem of label switching. Figure 2 displays marginal posterior distributions of the edge probabilities for M_1 and M_2 when data are generated by the M_2 model with $\beta' = (-1.0, -1.4, -0.5, 1.0)$. The precision of the posterior distributions increase in various degree when M_2 is assumed. Figure 2 also displays marginal posterior distributions of the edge probabilities by assuming M_2 (when M_1 is assumed, the marginal posterior distributions of the edge probabilities are more or less the same for both datasets) when data are generated by M_2 with $\beta' = 0$ which corresponds to $\theta_k(\mathbf{z}_i) = 1/3$, $k = 0, 1, 2$, $i = 1, \ldots, v$. The marginal posterior distributions under M_1 and M_2 are then more or less equal.

In Figure 3 the marginal posterior probability to belong to the correct class for M_2 is plotted against the marginal posterior probability to belong to the correct class for M_1 when $\beta' = (-1.0, -1.4, -0.5, 1.0)$. With one exception, the probabilities are located above the diagonal, suggesting that our proposed model substantially improves prediction of class affiliation when data are such that the posterior probabilities depend on known actor attributes.

5.2 A Data Example

We now apply the suggested approach to the countries trade network data described in Wasserman and Faust (1994), pp. 64–65. The data set includes five dichotomous and directional relations measured on a selection of 24 countries, and 4 attribute variables rejecting the economic and social characteristics of the countries. In our example we chose on of the relations, imports of food and live animals, and one of the attribute

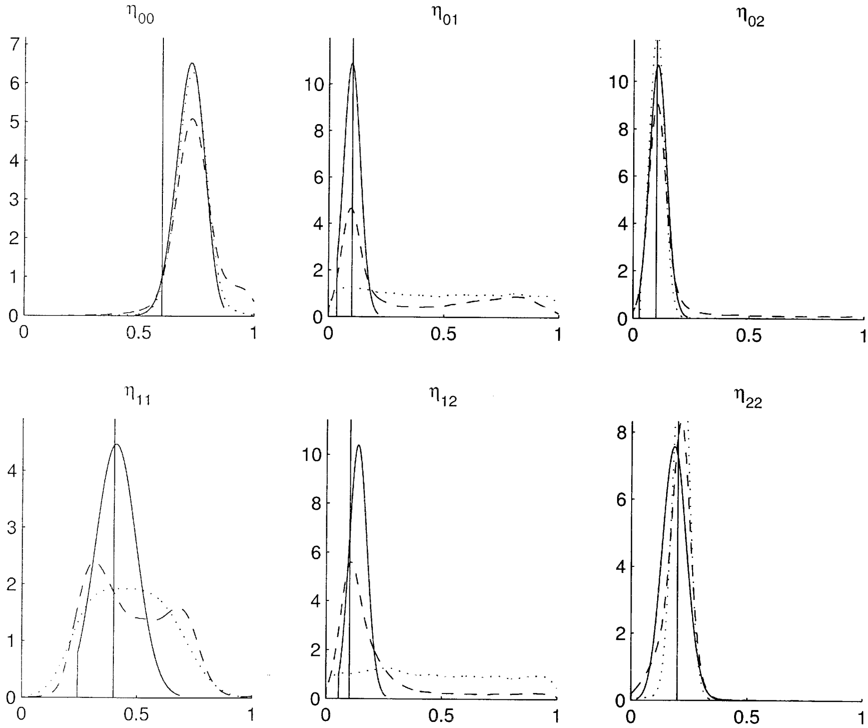

FIGURE 2 Marginal posterior distributions of η for M_2 for $\beta' = (-1, -1.4, -0.5, 1.0)$ (solid curves), $\beta' = \mathbf{0}$ (dotted curves), and M_1 (dashed curves). A solid line is drawn at the true parameter values.

variables, $p = 1$, energy consumption per capita in 1980. The countries are, with vertex labels between the first parenthesis and attribute values between the second parenthesis, Algeria (1)(814), Argentina (2)(2161), Brazil (3)(1101), China (4)(618), Czechoslovakia (5)(6847), Ecuador (6)(692), Egypt (7)(595), Ethiopia (8)(24), Finland (9)(6351), Honduras (10)(292), Indonesia (11)(266), Israel (12)(2813), Japan (13)(4649), Liberia (14)(502), Madagascar (15)(74), New Zealand (16)(4816), Pakistan (17)(224), Spain (18)(2944), Switzerland (19)(5223), Syria (20)(964), Thailnad (21)(370), United Kingdom (22)(5363), United States (23)(11626) and Yugoslavia (24)(2402). To reduce the number of probabilities, the relational variable is symmetrized as follows: if imports go both ways, then $x_{ij} = x_{ji} = 1$, else $x_{ij} = x_{ji} = 0$ even if import goes one way.

Since the number of classes are predetermined to three, $c = 3$, the vector of unknown regression coefficients is given by $\beta' = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})$. To ease the subjectivity in our analysis, we only entertain the relatively

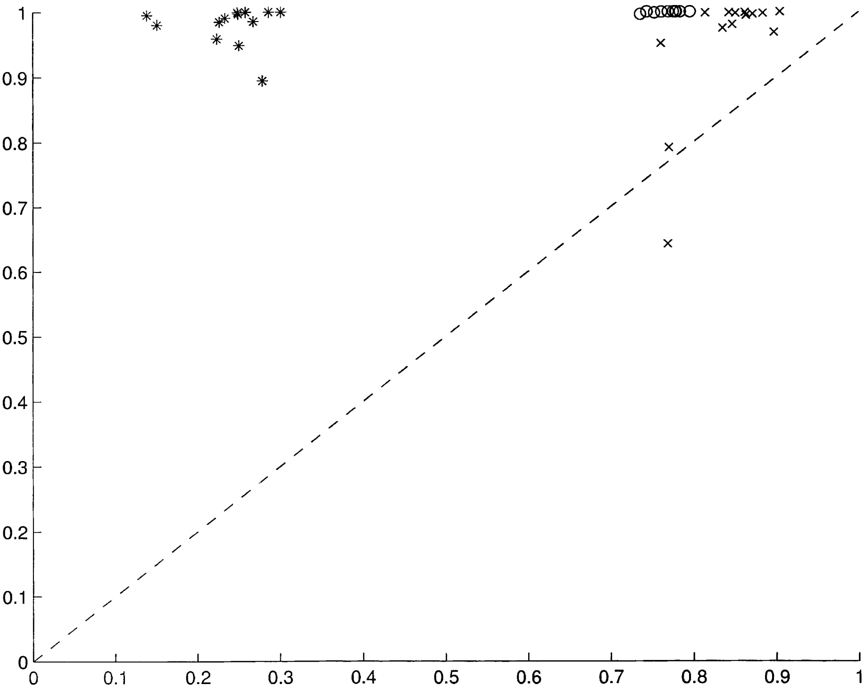

FIGURE 3 Marginal posterior probabilities to belong to the correct class for M_1 (x-axis) and for M_2 (y-axis). The vertices belonging to class 0, class 1 and class 2 are labeled with circles, stars and x, respectively.

weak informative prior on β and reference prior on $\boldsymbol{\eta}$ specified in the simulated data example. As in the previous example, we are forced to impose the same rather informative prior on Σ^{-1} due to the small size of the network in order to facilitate for the Gibbs sampler to converge. The marginal posterior distributions of the MNP model parameters are given in Figure 4. The corresponding posterior means and root mean square errors are given in Table 1. As in the simulated example, the marginal posterior distribution of Σ is centered on the identity matrix since the informative prior of Σ^{-1} is centered on the identity matrix. Furthermore, it is instructive to notice that the impact of the covariates on the class affiliation are limited due to small weights β_{11} and β_{21}. This result entails that the location and shape of the edge probabilities given in Figure 5 are rather similar for M_1 and M_2.

In Figure 6 the posterior class affiliation probabilities under M_2 are plotted against the posterior class affiliation probabilities under M_1. The

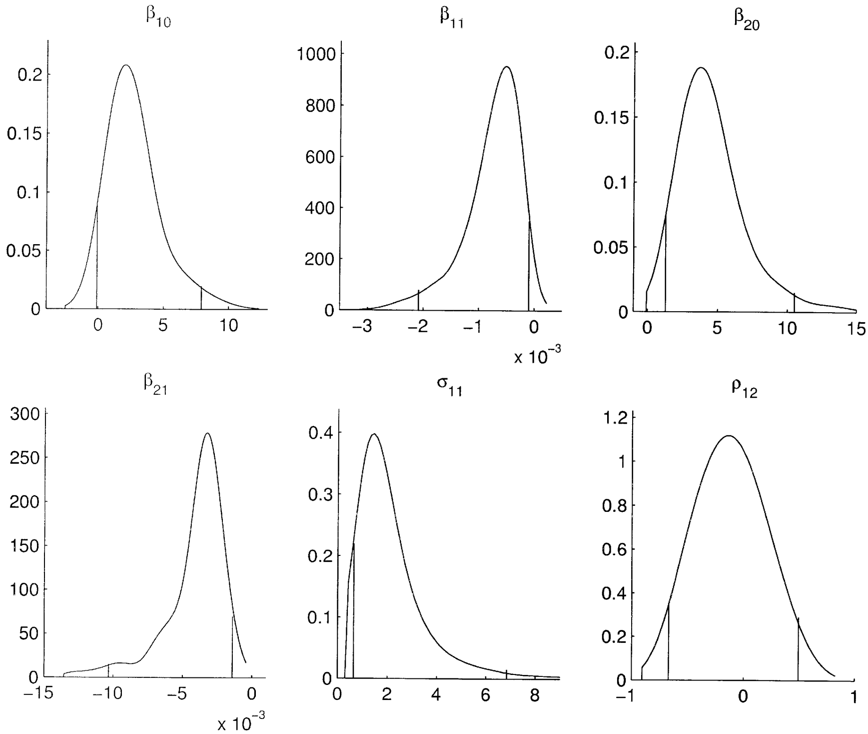

FIGURE 4 Posterior distributions of MNP model parameters for the entertained set of prior distributions. Solid vertical lines depict the lower and upper boundaries of 95% credibility intervals.

probability to belong to class 0, class 1 and class 2 is marked with a circle, a star and an x, respectively. Our prime concern is addressed to the actors whose class affiliation probabilities differ between the models,

TABLE 1 MCMC Results for the MNP Model Parameters

Parameter	Mean	RME
β_{10}	2.67	1.99
β_{11}	−0.0008	0.0005
β_{20}	4.52	2.34
β_{21}	−0.004	0.002
σ_{11}	2.23	2.07
σ_{22}	2.13	1.76
ρ_{12}	−0.12	0.31

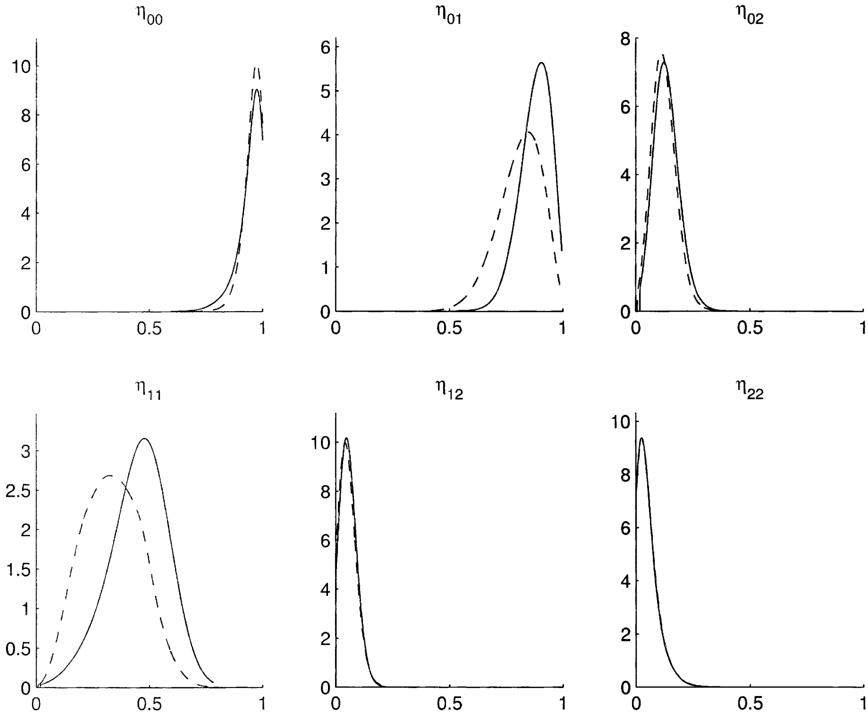

FIGURE 5 Posterior distributions of the edge probabilities for the entertained set of prior distributions for M_1 (solid curve) and M_2 (dashed curve).

i.e., are markedly located off the diagonal. Those five actors, China (4), Czechoslovakia (5), Indonesia (11), Thailand (21), and Yugoslavia (24), are labeled in Figure 6. By interpreting the MNP model parameters, we see that due to the negativity of β_{11} and β_{21} (although rather weak), and uncorrelated utilities (due to an informative prior), the probability to belong to class 0 increases for actor 5 due to relatively high covariate values and decreases for actors 4, 11, 21 and 24 due to relatively low covariate values. It is further interesting to note that the probability to belong to class 2 is more or less equal for both models.

The entropy provides a natural measure of uncertainty when a desirable property is that a dominating probability of a discrete random variable yields a low value, whereas equal values yields the maximum value. The entropy of a discrete random variable Y is defined by

$$H = -\Sigma p_y \log p_y$$

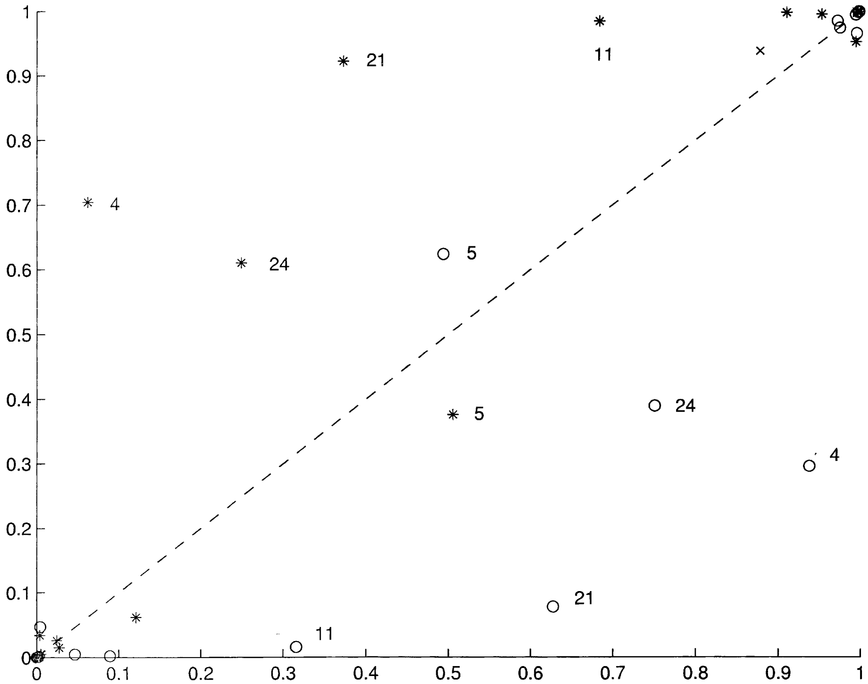

FIGURE 6 Posterior class affiliation probabilities for M_1 (x-axis), and M_2 (y-axis). Class 0, class 1 and class 2 is marked with a circle, a star and an x, respectively.

where $p_y = \Pr(Y = y)$, and the base of the logarithm is optional. A small value in H implies less uncertainty in the distribution of y. Since the class affiliation probabilities take three values, the normalized values $H/\log(3)$ ranges from zero to unity. The entropies of all actors under both models are listed in Table 2. The domination of one probability to belong to a specific class seems to increase when covariates are considered.

Our analysis shows roughly that Czechoslovakia, Finland, Japan, Spain, Switzerland, United Kingdom and United States are likely to belong to class 0; Argentina, Brazil, China, Egypt, Indonesia, Israel, New Zealand, Pakistan, Thailand and Yugoslavia are likely to belong to class 1, and Algeria, Ecuador, Ethiopia, Honduras, Liberia, Madagascar and Syria are likely to belong to class 2. A blockmodel analysis is performed by Wasserman and Faust (1994), pp. 403-406 on this data set, where they measured structural equivalence by using the Pearson product moment correlation coefficient on three relations: manufactured goods, raw materials, and diplomatic ties. All of these relations are directional and dichotomous. They identified

TABLE 2 Entropies of the Class Affiliation Probabilities under both Models.

Country	H_{M1}	H_{M2}	H_{M1}/H_{M2}
1	0.373	0.232	1.61
2	0.029	0.004	6.80
3	0.189	0.029	6.56
4	0.233	0.607	0.38
5	0.693	0.662	1.05
6	0.004	0.002	2.26
7	0.018	0.005	4.00
8	0.000	0.000	1.00
9	0.126	0.078	1.62
10	0.000	0.000	1.00
11	0.624	0.081	7.68
12	0.020	0.011	1.93
13	0.012	0.011	1.18
14	0.000	0.000	1.00
15	0.000	0.000	1.00
16	0.033	0.190	0.17
17	0.301	0.014	21.82
18	0.026	0.148	0.18
19	0.115	0.119	0.96
20	0.000	0.000	1.00
21	0.660	0.274	2.41
22	0.032	0.031	1.03
23	0.009	0.001	9.00
24	0.561	0.669	0.84

the following six positions (classes) by using complete link hierarchical clustering:

- V_0': Japan, United Kingdom, United States.
- V_1': China, Czechoslovakia, Indonesia, Spain, Yugoslavia.
- V_2': Argentina, Brazil, Finland, New Zealand, Pakistan, Switzerland, Thailand.
- V_3': Algeria, Egypt, Syria.
- V_4': Ecuador, Honduras, Israel.
- V_5': Ethiopia, Liberia, Madagascar.

The constellation of V_0 equals V_0' together with Czechoslovakia and Spain from V_1', and Finland and Switzerland from V_2'. The constellation of V_1 equals China, Indonesia and Yugoslavia from V_1', together with Argentina, Brazil, New Zealand, Pakistan and Thailand from V_2', Egypt from V_3' and Israel from V_4'. Finally, apart from Egypt and Israel, the constellation of V_2 equals V_3', V_4' and V_5'.

However, perphaps a 4-, 5- or 6-block solution would be more appropriate. This raises the relevant question: What is a reasonable number of classes? Our analysis is performed for a predetermined number of classes chosen quite arbitrarily. A proper analysis should devote some attention to assess the number of classes. The problem of comparing a collection of models that reflect hypothesis about the data is considered in Koskinen and Tallberg (2004). This is based on a method of computing model marginal likelihoods for Bayesian model comparisons from the output of Metropolis Hastings MCMC-chains, developed by Chib (1995) and Chib and Jeliazkov (2001).

6 DISCUSSION

In many applications, data are such that class affiliation to a large extent is governed by a priori actor attributes. Many approaches to blockmodels proposed ignore such information. In this paper, an extension to Nowicki and Snijders (2001) a posteriori blockmodels is presented for such data. The class affiliation probabilities are modeled conditional on the actor attributes via a MNP model. Computation of the posterior distributions of the model parameters and predictive posterior distributions of the class affiliation involves a Gibbs sampler which only requires draws from standard distributions.

Assumptions of independence and conditional independence between the units of analysis are common in social networks. In the introduced model, we assume that the probability distribution of the relation between two vertices depends on the class affiliation, and they in their turn depend on attribute values of the two vertices. By conditioning on the class affiliation of the vertices, the relations are independent. A challenge would be development of more elaborate probabilistic models that consider more complex conditional dependence assumptions. Computational obstacles have prevented such considerations in the past, but developments of computer intensive analysis methods in the last decade facilitate for such modeling. Frank and Strauss (1986) generalized the dyad independence models by introducing the notion of Markov dependency between dyads.

REFERENCES

Aitchison, J. & Bennett, J. A. (1970). Polychotomous quantal response by maximum indicant. *Biometrica*, *57*, 253–262.

Albert, J. & Chib, S. (1993). Bayesian analysis of binary and polychotomous data. *Journal of the American Statistical Association*, *88*, 669–679.

Chib, S. (1995). Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, *90*, 1313–1321.

Chib, S. & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, *96*, 270–281.

Fienberg, S. E. & Wasserman, S. (1981). Categorical data analysis of aingle sociometric relations. In S. Leinhardt (ed.), *Sociological Methodology-1981*, San Fransisco: Jossey-Bass, pp. 156–192.

Frank, O. (1988a). Triad count statistics. *Discrete Mathematics*, *72*, 141–149.

Frank, O. (1988b). Random sampling and social networks: A survey of various approaches, *Mathematiques. Informatique et Sciences Humaines*, *26*, 19–33.

Frank, O. & Harary, F. (1982). Cluster inference by using transitivity indices in empirical graphs *Journal of the American Statistical Association*, *77*, 835–840.

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

Gelfand, A. E. Hills, S. E. Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, *85*, 972–985.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints, *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*. Alexandria, VA: *American Statistical Association*, 571–578.

Gilks, W. R. Richardson, S. & Spiegelhalter, D. J. (1996). Introducing markov chain monte carlo. *In* W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice* (pp. 1–19). London: Chapman and Hall.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis, & *Journal of the American Statistical Association*, *97*, 1090–1098.

Holland, P. Laskey, K. B. & Leinhardt, S. (1983). Stochastic blockmodels: Some first steps. *Social Networks*, *5*, 109–137.

Holland, P. & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, *76*, 33–65.

Janson, S. & Nowicki, K. (1991). The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probability Theory and Related Fields*, *90*, 341–375.

Koop, G & Poirier, D. (1993). Bayesian analysis of logit models using natural conjugate priors, *Journal of Econometrics*, *56*, 323–340.

Koskinen, J. & Tallberg, C. (2004). A Bayesian approach to estimate the number of blocks in blockmodels. (Work in progress).

McCulloch, R; Polson, N. G. & Rossi, P. E. (200). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, *99*, 173–193.

McCulloch, R. & Rossi, P. E. (1994). An exact likelihood analysis of the multinominal probit model. *Journal of Econometrics*, *64*, 207–240.

Nowicki, K. & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, *96*, 1077–1087.

Richardson, S. & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society*, *59*, 731–792.

Snijders, T. A. B. & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent blockstrucutre. *Journal of classification*, *14*, 75–100.

Wang, Y. J. & Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, *82*, 8–19.

Wasserman. S & Anderson, C. (1987). Stochastic *a posteriori* block-models: Construction and assessment. *Social Networks*, *9*, 1–36.

Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*, New York: Cambridge University Press.

White, H. C., Boorman, S., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, *81*, 730–779.

Wong, G. Y. (1987). Bayesian models for directed graphs. *Journal of the American Statistical Association*, *82*, 140–148.