Research Report

Department of Statistics

No. 2003:4

A Bayesian Approach to Modeling Stochastic Blockstructures with Covariates

Christian Tallberg

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

A Bayesian Approach to Modeling Stochastic Blockstructures with Covariates

Christian Tallberg

Abstract

We consider the problem of partitioning members in random graphs with similar relational structure into subsets called blocks when the block labels are unobserved or latent. Most statistical research on this topic called blockmodels have been approached from a classical point of view. Recently, a Bayesian approach to blockmodels has been presented by Snijders and Nowicki (1997), and Nowicki and Snijders (2001), where the probability of a relation between two actors depends only on the blocks to which the actors belong but is independent of the actors. In this paper, we extend their model to include covariates on actor level, and the block affiliation probabilities are modeled conditional on the covariates via a multinomial probit model. Posterior distributions of the model parameters, and predictive posterior distributions of the block affiliation probabilities are computed by using a straight forward Gibbs sampling algorithm. The proposed model is illustrated on both real and simulated data.

Keywords: Bayesian analysis; Blockmodels; Gibbs sampling; Multinomial probit; Random graphs;.

1 Introduction

We are in our daily social environment involved in relationships of many sorts: economic, political, biological, sociological, criminological and affective, to mention a few. Relationships among social actors (which besides individuals can be organizations, countries etc.) and the patterns and implications of these relationships give rise to structures. The actors and their relational structure are usually represented by graphs, where the actors are referred to as *vertices* and the relations are referred to as *edges* (undirected graphs) or *arcs* (directed graphs).

Most methods in a field referred to as social network analysis are concerned with the description of network structural properties. One such formal property is *structural equivalence*. A definition is given in Lorrain and White (1971) which, briefly stated, says that two actors with identical attribute values, and thus with the same relational features are structural equivalent. We can then define a deterministic approach to blockmodels, first given by White, Boorman and Breiger (1976), as a partition of actors into discrete subsets, where actors in the same subset 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 relational 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 block labels are known, this approach is called a priori blockmodeling. When the block labels are unknown, Wasserman and Anderson (1987) proposed a blockmodeling procedure where block 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. For a more basic clarification on blockmodels and other social network concepts, the reader is referred to Wasserman and Faust (1994).

Most statistical research on blockmodels has been done from a classical point of view. During recent years, the interest of developing methods within the Bayesian paradigm has increased. In a paper by Snijders and Nowicki (1997), a Bayesian approach to blockmodeling of graphs is presented, where the number of blocks is restricted to two, and the probability of an edge between two actors depends only on the blocks to which the actors belong but is 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.

The motivation for this paper is that the a posteriori blockmodels proposed in those articles completely ignore available information on actor level. In their settings the block affiliation probabilities are the same for all actors. In this paper, we extend the model outlined by Nowicki and Snijders (2001) to include observable covariates on actor level, and let the block affiliation probabilities depend on those covariate values. Using available information from covariates on actor level, should improve the prediction of the block affiliation of each actor.

In our proposed model, the number of blocks is predetermined and fixed. The relationship between the block affiliation probabilities and the covariates is modeled with a probit model. Using the simulation based approach for the multinomial probit 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 block 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 multinomial probit 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

To a certain extent we will follow the notation of Nowicki and Snijders (2001). Let $\mathcal{V} = \{1, ..., n\}$ be the vertex set, where n is the known order of the graph and $\mathcal{N}$ the set of all distinct ordered pairs of vertices (i, j). We assume a general relational structure on this set of vertices which is represented by its adjacency matrix $\mathbf{x} = (x_{ij}), (i, j) \in \mathcal{N}$, where the element x_{ij} is an observed value of a relation from vertex i to vertex j. Let $x_{ij} \in \mathcal{R}$, where $\mathcal{R} = \{0, 1, ..., R-1\}$ is the range space, i.e. the set of possible values of a relation from vertex i to vertex j. Self-loops are not allowed which implies that $x_{ij} = 0$ for $i = j$. In the special cases of graphs and digraphs we have that $\mathcal{R} = \{0, 1\}$. Pairwise relations from i to j and from j to i, takes values in the set $\mathcal{R}^2 = \{\mathbf{r} = (r_u, r_v) \,| r_u, r_v \in \mathcal{R}\}$. Since it is not necessary that all elements of $\mathcal{R}^2$ can occur, we define $\mathcal{S}$ as a subset of $\mathcal{R}^2$ containing potential values of pairwise relations. Thus, for every pair $(i, j) \in \mathcal{N}$, there is an $(r_u, r_v) \in \mathcal{S}$ such that the relation from i to j is r_u and the relation from j to i is r_v. Symmetry assumptions about the pairwise relations implies that for each $(r_u, r_v) \in \mathcal{S}$ it holds that $(r_v, r_u) \in \mathcal{S}$.

We assume that $\mathcal{V}$ is partitioned into b mutually exclusive non-empty vertex subsets called blocks denoted by $\mathcal{B} = \{0, ..., b-1\}$. Let $y_i \in \mathcal{B}$ be a block label where $y_i = k$ if vertex i belongs to block k, $i \in \mathcal{V}$ and $k \in \mathcal{B}$.

The dyad involving i and j is characterized by $(x_{ij}, x_{ji}, y_i, y_j, \mathbf{z}_i, \mathbf{z}_j)$, where $\mathbf{z}_i' = (z_{i1}, ..., z_{ip})$ and $\mathbf{z}_j' = (z_{j1}, ..., z_{jp})$ are vectors of known covariate values of actors i and j, respectively. Conditional on the block labels (y_i, y_j), $(i, j) \in \mathcal{N}$, the pairs (x_{ij}, x_{ji}), $(i, j) \in \mathcal{N}$, are assumed to be independent random vectors with distribution

$$\begin{aligned} & \Pr\left[(x_{ij}, x_{ji}) = \mathbf{r} \,| y_1, ..., y_n\right] \\ = & \Pr\left[(x_{ij}, x_{ji}) = \mathbf{r} \,| y_i, y_j\right] = \eta_{\mathbf{r}}\left(y_i, y_j\right), \end{aligned}$$

where the array

$$\boldsymbol{\eta} = \eta_{\mathbf{r}}(k, h) \ \text{ for } \mathbf{r} \in \mathcal{S},\ k, h \in \mathcal{B}$$

of block dependent dyad probabilities satisfies the restriction

$$\sum_{\mathbf{r} \in \mathcal{S}} \eta_{\mathbf{r}}(k, h) = 1 \text{ for all } k, h \in \mathcal{B}.$$

Due to the assumed symmetry in the relations, the restriction

$$\eta_{(r_u, r_v)}(k, h) = \eta_{(r_v, r_u)}(h, k)$$

holds and the model is over-parametrized. This is easily remedied by removing the excess parameters from $\boldsymbol{\eta}$ so that the elements of $\boldsymbol{\eta}$ are $\eta_{(r_u, r_v)}(k, h) = \eta_{\mathbf{r}}(k, h)$ for $(u, v) \in \mathcal{S}$, $k \leqslant h$.

To define our stochastic blockmodel, we assume that the block labels are unobserved (latent) iid where the probability that vertex i belongs to block y_i is given by

$$\Pr(y_i \,| \mathbf{z}_i) = \theta(y_i \,| \mathbf{z}_i).$$

for block $y_i \in \mathcal{B}$. Hence, the joint distribution of the $1 \times n$ vector of unknown block labels $\mathbf{y} = (y_1, ..., y_n)$ is given by

$$\Pr(y_1, ..., y_n | \mathbf{z}_1, ..., \mathbf{z}_n) = \prod_{i=1}^{n} \Pr(y_i | \mathbf{z}_i) = \prod_{i=1}^{n} \theta(y_i | \mathbf{z}_i).$$

Since the conditional distribution of relations $\mathbf{x}$ given $\mathbf{y}$ and $\boldsymbol{\eta}$ is given by

$$\Pr(\mathbf{x} | \mathbf{y}, \boldsymbol{\eta}) = \prod_{\mathbf{r} \in \mathcal{S}} \prod_{0 \leqslant k \leqslant h \leqslant b-1} \{\eta_{\mathbf{r}}(k,h)\}^{e_{\mathbf{r}}(k,h)},$$

where $e_{\mathbf{r}}(k,h)$ are the edge frequencies between block k and block h, the stochastic blockmodel given by the joint distribution of $(\mathbf{x}, \mathbf{y})$, can be written as

$$\begin{aligned} &\Pr(\mathbf{x}, \mathbf{y} | \boldsymbol{\eta}, \theta(y_1 | \mathbf{z}_1), ... \theta(y_n | \mathbf{z}_n), \mathbf{z}_1, ..., \mathbf{z}_n) \\ &= \left(\prod_{i=1}^{n} \theta(y_i | \mathbf{z}_i) \right) \left(\prod_{\mathbf{r} \in \mathcal{S}} \prod_{0 \leqslant k \leqslant h \leqslant b-1} \{\eta_{\mathbf{r}}(k,h)\}^{e_{\mathbf{r}}(k,h)} \right). \end{aligned} \tag{1}$$

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

To ease the notation we will denote the probability that vertex i belongs to block k with $\theta_k(\mathbf{z}_i)$.

The model given by (1), which is an extension of the model presented by Nowicki and Snijders (2001), includes observable covariates on actor level, such as gender, age, income and so on. The block affiliation probabilities depend on those covariate values which should improve the prediction of $\mathbf{y}$. Thus, unlike the simpler model of Nowicki and Snijders (2001), where the probability to belong to block k, θ_k, is the same for all actors, our proposed model is richer since it allows the block affiliation probabilities, $\theta_k(z_i)$, to vary between the actors.

The relationship between the block affiliation probabilities and the covariates is modeled with the multinomial probit model introduced by Aitchison and Bennet (1970). In the special case of two blocks, the multinomial probit model reduces to the binary probit model. A brief discussion of the binary probit model and its generalized version is provided in the next section.

3 The multinomial probit model

We open this section with a short review of the special case $b = 2$, which in the graph setting means that we have two blocks, block 0 and block 1. Then y_i is Bernoulli distributed with probability $\theta_1(\mathbf{z}_i)$ that actor i belongs to block 1. Assuming a binary probit model implies that the $\theta_1(\mathbf{z}_i)$ are related to the set of covariates through

$$\theta_1(\mathbf{z}_i) = \Pr(y_i = 1 \,|\mathbf{z}_i, \boldsymbol{\beta}) = \Phi(\mathbf{z}_i'\boldsymbol{\beta}), \tag{2}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and Φ is the standardized normal distribution function. The posterior density of $\boldsymbol{\beta}$ is largely intractable but Albert and Chib (1993) presented a simulation-based approach for computing the posterior of $\boldsymbol{\beta}$. They considered a standard latent variable interpretation of the probit model. Thus, (2) may be written in terms of n independent latent variables $W_1, ..., W_n$, where W_i is $N(\mathbf{z}_i'\boldsymbol{\beta},1)$ distributed, by defining $y_i = 1$ if $W_i > 0$ and $y_i = 0$ if $W_i \leqslant 0$. Given y_i, the distribution of W_i follows a truncated normal distribution, and therefore simulating from the full conditional posterior distributions of $\boldsymbol{\beta}$ and W_i is easy by using Gibbs sampling. Note that in our analysis we are not concerned with the characteristics of W_i. They are only introduced to facilitate the computation of the posterior of $\boldsymbol{\beta}$.

We now proceed to briefly review the generalized case with $b > 2$. Each y_i is now a multiple choice block label. Like in the two choice case, the key idea is to introduce a latent vector, here given by $\mathbf{V}_i' = \left(V_{i0}, V_{i1}, ..., V_{i(b-1)}\right)$, for each actor. Let $\mathbf{u}_i' = (\mathbf{u}_{i1}, ..., \mathbf{u}_{ip})$ be a $b \times p$ matrix of known covariates for actor i, and $\boldsymbol{\beta}$ a $p \times 1$ vector of unknown parameters. The structural equations of the multinomial probit model are then given by

$$\mathbf{V}_i = \mathbf{u}_i'\boldsymbol{\beta} + \xi_i,$$

where each ξ_i is $N_b(0, \Psi)$ and Ψ is of dimension $b \times b$. Each y_i is then a function of $\mathbf{V}_i$ given by

$$y_i(\mathbf{V}_i) = \arg\max_{k \in \mathcal{B}} V_{ik}.$$

We shall follow the procedure of McCulloch and Rossi (1994), and express each v_{ik} relative to v_{i0} in terms of $w_{ik} = v_{ik} - v_{i0}$. By letting $\mathbf{W}_i' =$

$\left(W_{i1}, W_{i2}, ..., W_{i(b-1)}\right)$, the $(b-1)$ dimensional latent variable can be expressed as

$$\mathbf{W}_i = \mathbf{z}_i'\boldsymbol{\beta} + \varepsilon_i, \tag{3}$$

where each ε_i is $N(0, \Sigma)$. The covariance matrix Σ is of dimension $(b-1) \times (b-1)$, and $\mathbf{z}_i'$ is a $(b-1) \times p$ matrix transformed from the original matrix $\mathbf{u}_i'$ by subtracting the first row from the last $(b-1)$ rows. Each y_i is then a function of $\mathbf{W}_i$ given by

$$y_i(\mathbf{W}_i) = \begin{cases} 0 & \text{if } \max\limits_k W_{ik} \leqslant 0 \\ \arg\max\limits_k W_{ik} & \text{if } \max\limits_k W_{ik} > 0. \end{cases}$$

Thus, $y_i = 0$ if all the w_{ik} are non-positive, otherwise y_i equals the index of the biggest positive w_{ik}. Since $\mathbf{W}_i$ is a continuous random vector, the probability that at least two elements are equal is zero.

Besides the block labels y_i, the random graph setting includes an unknown set of relational probabilities $\boldsymbol{\eta}$. Our immediate concern is therefore the computation of the posterior distributions of $y_i, \boldsymbol{\eta}$, $\boldsymbol{\beta}$ and Σ. As mentioned earlier, the introduction of $\mathbf{W}_i$ is only to facilitate the computation of the posterior distributions of the probit parameters $\boldsymbol{\beta}$ and Σ.

The conditional distribution, now given by $y_i\,|\mathbf{z}_i, \mathbf{x}, \boldsymbol{\beta}, \Sigma, \boldsymbol{\eta}$, depends on data and the unknown model parameters $\boldsymbol{\beta}$, Σ and $\boldsymbol{\eta}$. The parameters $\boldsymbol{\beta}$ and Σ are not fully identified, since the distribution of $y_i(\mathbf{W}_i)$ equals the distribution of $y_i(a\mathbf{W}_i)$ for all $a > 0$. This means that given an observation, the likelihood is such that $L(\boldsymbol{\beta}, \Sigma, \boldsymbol{\eta}) = L(a\boldsymbol{\beta}, a^2\Sigma, \boldsymbol{\eta})$. One way to solve the identification problem is to assign a value to one of the parameters. Since we are only making relative comparisons, one of the diagonal elements in Σ may be normalized to unity. In this paper we adopt the approach described in detail in McCulloch et al (2000), where the first element in the covariance matrix, σ_{11}, is set to 1. Thus, $\mathbf{W}_i$ is a latent vector with a truncated multivariate normal distribution $N(\mathbf{z}_i'\boldsymbol{\beta}, \Sigma\,|\sigma_{11} = 1)$. It is now easy to see how the multinomial probit model reduces to the binary probit model, and the probability (2) arrives at $\Phi(\mathbf{z}_i'\beta)$.

The main obstacle in implementation of the multinomial probit model has been to compute 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 multinomial probit model feasible. An alternative to the multinomial probit model would be the multinomial logit model with uncorrelated choices; see Koop and Poirier (1993).

4 Prior assignments and computation of posterior distributions

A Bayesian analysis requires the specification of a prior over the parameters $(\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \Sigma)$, and the computation of the posterior distribution given by

$$p\left(\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \Sigma \,|\mathbf{x}, \mathbf{z}\,\right) \propto p\left(\mathbf{x}, \mathbf{z}\,|\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \Sigma\,\right) p\left(\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \Sigma\right) \tag{4}$$

according to Bayes theorem, where $\mathbf{z}$ is a $p \times n$ matrix of known covariate values. Since mutual independence between $\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\beta}$ and Σ is a realistic assumption here, the prior distribution can be written as

$$p\left(\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \Sigma\right) = p\left(\mathbf{y}\right) p\left(\boldsymbol{\eta}\right) p\left(\boldsymbol{\beta}\right) p\left(\Sigma\right).$$

The posterior joint distribution given by (4) is highly intractable, but by using conjugate priors on the parameters, the full conditional posterior distribution of each parameter is possible to compute. By implementing the Gibbs sampler, which is a computer-intensive method based on properties of Markov chains, we are provided with an alternative method that allows us to generate random samples from the marginal distributions indirectly by iteratively sampling from the full conditional distributions.

A short review of the Gibbs sampling technique is as follows. Specify starting values for the parameters of interest. Then updated values of the parameters are obtained iteratively by alternately generating values from the conditional distributions. As the number of iterations approaches infinity, the Gibbs sampler generates accurate samples from the marginal distributions. For details and further references, see Gilks, Richardson and Spiegelhalter (1996).

To implement the algorithm, the following four full conditional posterior distributions are required:

- $p\left(\mathbf{y}\,|\boldsymbol{\eta}, \boldsymbol{\beta}, \Sigma, \mathbf{x}, \mathbf{z}\,\right)$
- $p\left(\boldsymbol{\eta}\,|\mathbf{y}, \boldsymbol{\beta}, \Sigma, \mathbf{x}, \mathbf{z}\,\right)$
- $p\left(\boldsymbol{\beta}\,|\mathbf{y}, \boldsymbol{\eta}, \Sigma, \mathbf{x}, \mathbf{z}\,\right)$
- $p\left(\Sigma\,|\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{x}, \mathbf{z}\,\right)$

4.1 Full conditional posterior of $\mathbf{y}$

By following the notation of Nowicki and Snijders (2001), we first define $d_{\mathbf{r}}(i,k)$ to be the number of relations $\mathbf{r} \in \mathcal{S}$, between $i \in \mathcal{V}$ and all $j \in \mathcal{V}$ that belongs to block $k \in \mathcal{B}$, which can be expressed more formally as

$$d_{\mathbf{r}}(i,k) = \sum_{j:(i,j)\in\mathcal{N}} I\{x_{ij} = \mathbf{r}\}\, I\{y_j = k\}\,.$$

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

$$\Pr\left(y_i = k \left| \{\mathbf{y}_j\}_{j\neq i}, \boldsymbol{\eta}, \boldsymbol{\beta}, \Sigma, \mathbf{x}, \mathbf{z}\right.\right) \propto \theta_k(\mathbf{z}_i) \prod_{\mathbf{r}\in\mathcal{S}} \prod_{h=0}^{b-1} \{\eta_{\mathbf{r}}(k,h)\}^{d_{\mathbf{r}}(i,k)}.$$

4.2 Full conditional posterior of $\boldsymbol{\eta}$

To determine the full conditional posterior distribution of each vector $\eta_{\mathbf{r}}(k,h)$, we first note that the edge frequencies is multinomially distributed data. If the prior distribution of $\eta_{\mathbf{r}}(k,h)$ is conjugate Dirichlet with parameters $a_{\mathbf{r}}(k,h)$, the full conditional posterior distribution of each $\eta_{\mathbf{r}}(k,h)$ will have the form

$$\begin{aligned} p\{\eta_{\mathbf{r}}(k,h)\,|\mathbf{y}, \boldsymbol{\beta}, \Sigma, \mathbf{x}, \mathbf{z}\} &= p\{\eta_{\mathbf{r}}(k,h)\,|\mathbf{y}, \mathbf{x}\} \\ &\propto p\{\mathbf{y}, \mathbf{x}\,|\eta_{\mathbf{r}}(k,h)\}p\{\eta_{\mathbf{r}}(k,h)\} \\ &= p\{\mathbf{x}\,|\mathbf{y}, \eta_{\mathbf{r}}(k,h)\}p\{\mathbf{y}\,|\eta(k,h)\}p\{\eta_{\mathbf{r}}(k,h)\} \\ &\propto \prod_{\mathbf{r}\in\mathcal{S}} \{\eta_{\mathbf{r}}(k,h)\}^{e_{\mathbf{r}}(k,h)+a_{\mathbf{r}}(k,h)-1}, \end{aligned} \tag{5}$$

which is a Dirichlet distribution with parameters $e_{\mathbf{r}}(k,h) + a_{\mathbf{r}}(k,h)\,, 0 \leqslant k \leqslant h \leqslant b-1$.

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 suggests that identifiability restrictions are imposed on the elements in $\boldsymbol{\eta}$.

4.3 Full conditional posterior of $\mathbf{W}_i$

To be able to compute the exact posterior distributions of $\boldsymbol{\beta}$ and Σ, we introduce n independent latent vectors $\mathbf{W}_i$ as discussed in Section 3. The full conditional posterior of each $\mathbf{W}_i$ is equal to the $N_{b-1}\left(\mathbf{z}_i'\boldsymbol{\beta}, \Sigma \,|\sigma_{11} = 1\right)$ distribution, $i = 1, ..., n$, truncated to the region

$$W_i \in \mathcal{R}^{p-1} : \max_k W_{ik} \leqslant 0,$$

if $y_i = 0$, and

$$W_i \in \mathcal{R}^{p-1} : \arg\max_k W_{ik} = y_i,$$

otherwise. There are various algorithms suggested that can draw a sample of $\mathbf{W}_i$ from this truncated distribution. A simple algorithm is to perform repetitive draws of $\mathbf{W}_i$ from $N_{b-1}\left(\mathbf{z}_i'\beta, \Sigma\right)$ until the condition is satisfied (reject/accept algorithm). A more effective alternative, outlined in McCulloch and Rossi (1994), is to draw each element, W_{ik}, from a truncated univariate normal distribution. For a detailed description of simulation from a truncated multivariate normal distribution; see also Geweke (1991). As noted in Section 3, the draws of $\mathbf{W}_i$ are of no interest per se and need not be saved after the termination of the iteration.

4.4 Full conditional posterior of $\boldsymbol{\beta}$

By selecting a proper conjugate $N\left(\boldsymbol{\beta}^*, \mathbf{B}^*\right)$ prior for $\boldsymbol{\beta}$, we have according to Bayes theorem that the full conditional posterior of $\boldsymbol{\beta}$ is $N\left(\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{B}}\right)$, where the mean and covariance matrix are given by

$$\tilde{\boldsymbol{\beta}} = \left(\mathbf{B}^{*-1} + \mathbf{z}'\Sigma^{-1}\mathbf{z}\right)^{-1}\left(\mathbf{B}^{*-1}\boldsymbol{\beta}^* + \mathbf{z}'\mathbf{W}\right)$$

and

$$\tilde{\mathbf{B}} = \left(\mathbf{B}^{*-1} + \mathbf{z}'\Sigma^{-1}\mathbf{z}\right)^{-1},$$

respectively, where $\mathbf{z} = (\mathbf{z}_1, ..., \mathbf{z}_n)'$ is a $n \times p$ matrix of known covariates.

4.5 Full conditional posterior of Σ

Due to various solutions of the identification problem that arises in the multinomial probit model, there are different approaches how to set priors on Σ and derive full conditional posteriors, see for example McCulloch and Rossi (1994). We shall adopt the approach suggested by McCulloch et al (2000), where it is suitable to reparametrize Σ first. As mentioned in Section 3, the first diagonal element in Σ is set to 1. First we denote ε_i in Equation (3) by $\left(\varepsilon_{i1}, ..., \varepsilon_{i(b-1)}\right)'$. Then we let $U = \varepsilon_{i1}$ and $V = \left(\varepsilon_{i2}, ..., \varepsilon_{i(b-1)}\right)'$, i.e. $\varepsilon_i' = (U, V)$. By letting $\gamma' = E(UV')$ and $\phi = E(VV') - \gamma\gamma'$, we can rewrite Σ as

$$\Sigma = E\left(\varepsilon_i \varepsilon_i'\right) = \left\{ \begin{array}{cc} E\left(U^2\right) & E\left(UV'\right) \\ E\left(VU\right) & E\left(VV'\right) \end{array} \right\} = \left\{ \begin{array}{cc} 1 & \gamma' \\ \gamma & \phi + \gamma\gamma' \end{array} \right\}.$$

Now, by choosing the following priors

$$\begin{aligned} \gamma &\sim N\left(\bar{\gamma}, C\right) \\ \phi^{-1} &\sim Wishart\left(m, D\right), \end{aligned}$$

we will obtain the following conditional posteriors

$$\gamma \sim N\{A_\gamma[\mathrm{vec}\left(\phi^{-1}V'U\right) + C^{-1}\bar{\gamma}], A_\gamma\} \quad (6)$$

$$\phi^{-1} \sim Wishart\{m + n, [D + (V - U\gamma')'(V - U\gamma')]^{-1}\}, \quad (7)$$

where $A_\gamma = \left(U'U\phi^{-1} + C^{-1}\right)^{-1}$, and vec is the vec operator. The vec operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. We parametrize the Wishart distribution so that $E\left(\phi^{-1}\right) = mD$.

Note that in each iteration of the Gibbs sampler, we have to perform an embedded Gibbs sampling in order to obtain the posterior distribution of the probit parameters from which we sample the last updated values of $\boldsymbol{\beta}$,γ and ϕ^{-1}. 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. For simplicity, we will in the given examples consider posterior blockmodeling for undirected graphs. Then the set containing potential relational values of pairwise relations is reduced to $\mathcal{S} = \{(0,0), (1,1)\}$, and $\eta_{\mathbf{r}}(k,h)$ is denoted by $\eta(k,h)$, $k,h = 0,1, k \leqslant h$. Since the relation between any two vertices now is a binary variable, the multinomial data model and its conjugated Dirichlet prior reduces to the binomial data model and its conjugated beta prior. In accordance with (5), the posterior distribution of each $\eta(k,h)$ is given by

$$\begin{aligned} & p(\eta(k,h) \,|\mathbf{x},\mathbf{y},\mathbf{z},\boldsymbol{\beta}) = p(\eta(k,h) \,|\mathbf{x},\mathbf{y}) \\ \propto \; & \{\eta(k,h)\}^{e_{(1,1)}(k,h)+a_{(1,1)}(k,h)-1}\{1-\eta(k,h)\}^{e_{(0,0)}(k,h)+a_{(0,0)}(k,h)-1}, \end{aligned}$$

which is a beta distribution with parameters $e_{(1,1)}(k,h) + a_{(1,1)}(k,h)$ and $e_{(0,0)}(k,h) + a_{(0,0)}(k,h)$.

In all analysis, we have used Markov chains of length 10,000 after a burn-in period of 20,000 observations. The numerical computation package Matlab was used for all computations on a standard PC.

As previously mentioned, we have predetermined the number of blocks. The simulated data set is generated so that the graph is composed of two blocks, whereas the real data set is analyzed under the assumption that the number of blocks is two and three.

5.1 Simulated example with two blocks

In this example we will use a model including one predictor and no intercept. Data was simulated as follows. A vector-valued observation of predictor values $\mathbf{z}$ was generated where each element was drawn iid from a uniform distribution on the interval $(-0.5, 0.5)$. The value of the single β-coefficient equals -2. The number of actors was determined to $n = 30$ of which 19 were allocated to block 0 and 11 to block 1 through the following procedure. Given the values of $\mathbf{z}$ and β, a vector of 30 observations was generated, where each observation was drawn iid from a $N(z_i\beta, 1)$-distribution. If element $i, i = 1, ..., 30$, was positive, actor i belonged to block 1, whereas

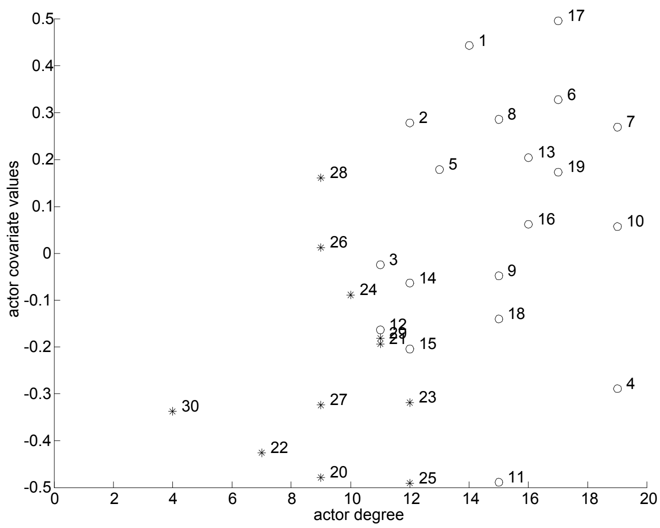

Figure 1: Plot of the actor degrees against actor covariate values for the simulated data set. Actors affiliated to block 0 and block 1 are represented with circles and stars, respectively.

if element i was negative, actor i belonged to block 0. Of the 30 generated observations constituting our population, 19 were negative and 11 were positive. Given the vector of block labels $\mathbf{y}$ and a vector of probabilities $\boldsymbol{\eta} = (\eta(0,0), \eta(0,1), \eta(1,1)) = (0.7, 0.3, 0.5)$, an adjacency matrix of order 30 was generated where the elements were drawn independently from Bernoulli distributions with probability $\eta(k,h)$. The columns and rows in the adjacency matrix were rearranged so that the first 19 actors belonged to block 0 and the last 11 belonged to block 1. Figure 1 shows the actor covariate values plotted against the actor degree.

To proceed with the Bayesian approach, we need to specify prior distributions of the model parameters. With two blocks, Σ is a scalar that takes the value 1 with probability 1, and therefore we are only concerned to determine priors on β and $\boldsymbol{\eta}$. As mentioned in Section 4.4, we consider normally distributed priors on β. They are all centered on 0 with various precisions expressing various degree of uncertainty about β.

For $\boldsymbol{\eta}$ we consider the following four sets of beta priors, representing different initial beliefs:

1. $\eta(0,0) \sim \text{beta}(15,5)$, $\eta(0,1) \sim \text{beta}(3,3)$ and $\eta(1,1) \sim \text{beta}(1,10)$.

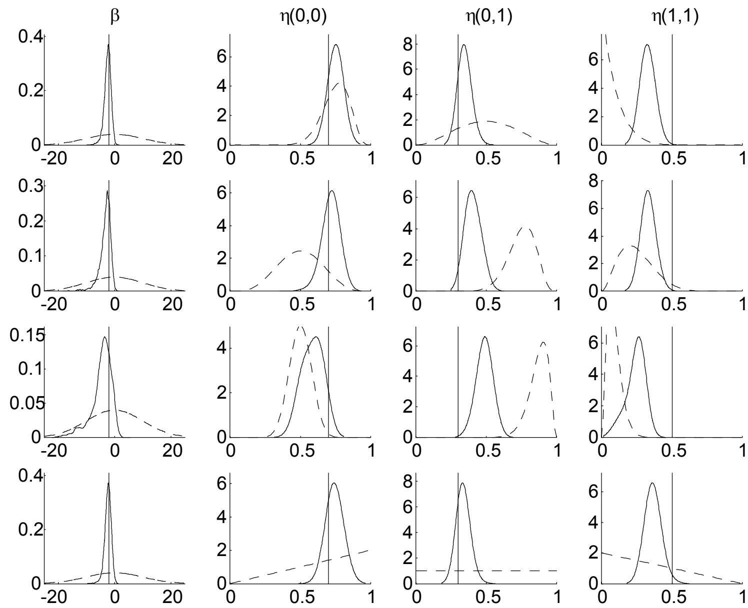

Figure 2: Prior (dashed curves) and posterior (solid curves) distributions of β and $\boldsymbol{\eta}$. The prior distribution of β is $N(0, 100)$ in all cases, whereas the sets of priors on $\boldsymbol{\eta}$ are prior 1 (top row), prior 2 (second row), prior 3 (third row) and prior 4 (bottom row). A solid line is drawn at the true parameter value.

2. $\eta(0,0) \sim$ beta $(5,5)$, $\eta(0,1) \sim$ beta $(15,5)$ and $\eta(1,1) \sim$ beta $(3,9)$.

3. $\eta(0,0) \sim$ beta $(20,20)$, $\eta(0,1) \sim$ beta $(20,3)$ and $\eta(1,1) \sim$ beta $(3,30)$.

4. A uniform distribution over the restricted domain $0 < \eta(1,1) < \eta(0,0) < 1$. Note that the problem of label switching discussed in the previous section is avoided by restricting the domain of $\boldsymbol{\eta}$.

The prior distributions and computed posterior distributions are displayed in Figure 2. We see that the posterior distributions are rather insensitive to the choices of priors, since they are quite similar although the priors of $\boldsymbol{\eta}$ are not.

Before we continue the analysis, we note that in order to obtain valid inference of the parameters of interest, the Gibbs sampler must converge

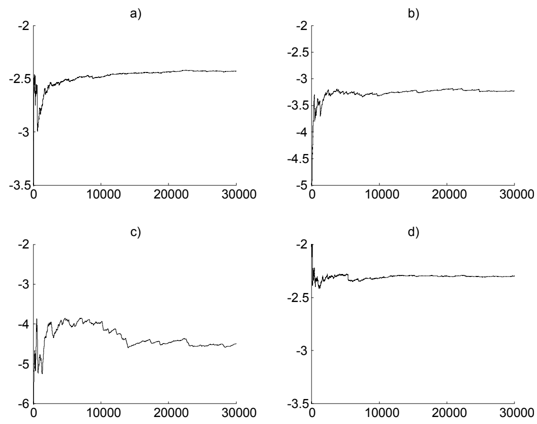

Figure 3: Posterior mean evolution of β under a) prior 1; b) prior 2; c) prior 3; d) prior 4 on $\boldsymbol{\eta}$, and a $N(0, 100)$ distributed prior on β.

which is the case according to the posterior means of β plotted in Figure 3.

The posterior probability to belong to the correct block for the model with covariates is plotted against the posterior probability to belong to the correct block for the model without covariates in Figure 4. Our immediate concern are the actors located off the diagonal, since their probabilities to belong to the correct block differs between the models. Since actors 1, 2, 5, 11, 12, 14, 15, 16 and 18 belong to block 0, it follows from the negativity of β that the posterior probabilities to belong to the correct block increase for actors 1, 2, 5 and 16 due to relatively high covariate values, and decrease for actors 11, 12, 14, 15 and 18 due to relatively low covariate values under prior 4. Analogously, it follows from the negativity of β that the posterior probabilities increase for actors 2, 5, 8, 13, 14, 19, 23, 24, 25, 27 and 29 due to "right" covariate values, whereas they decrease for actors 4, 9, 11, 12, 18, 26 and 28 due to "wrong" covariate values under prior 3. The obtained results seem to confirm that using available covariate information effectively via a probit model can improve prediction of block affiliation.

Figure 5 displays posterior distributions of β under four prior precisions, and prior 1 on $\boldsymbol{\eta}$. Although it looks as though the posterior means are robust to the choices of priors, except under the extremely informative prior where $Var(\beta) = 0.01$, we see that the root mean square error decreases with Bayesian methods if conceivable prior information on the parameters of

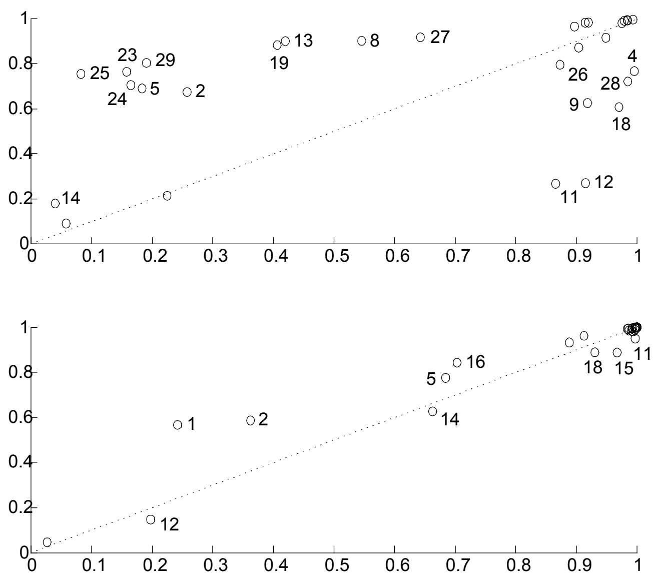

Figure 4: Posterior probabilities to belong to the correct block without covariates (x-axis) and with covariates (y-axis), with prior 3 (top row) and prior 4 (bottom row) on $\boldsymbol{\eta}$, and a $N(0, 100)$ distributed prior on β.

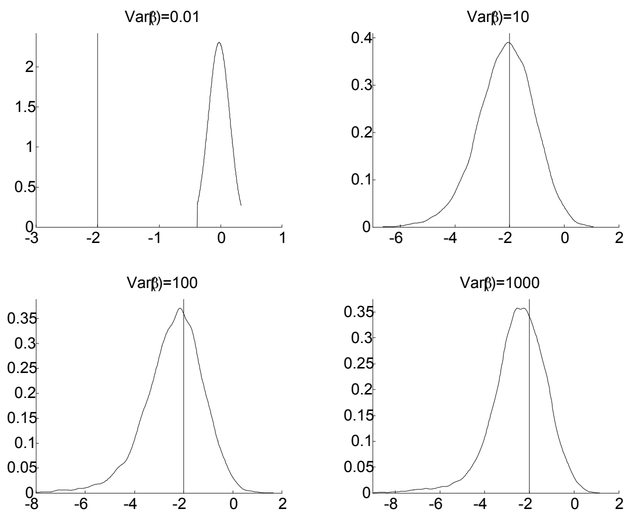

Figure 5: Posterior distributions of β computed on the simulated data set under normal distributed priors with mean 0 and various variances, and prior 1 on $\boldsymbol{\eta}$. A solid line is drawn at the true parameter value.

interest is implemented.

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 twenty-four countries, and four attribute variables reflecting the economic and social characteristics of the countries. The countries are, with vertex labels between parenthesis, Algeria (1), Argentina (2), Brazil (3), China (4), Czechoslovakia (5), Ecuador (6), Egypt (7), Ethiopia (8), Finland (9), Honduras (10), Indonesia (11), Israel (12), Japan (13), Liberia (14), Madagascar (15), New Zealand (16), Pakistan (17), Spain (18), Switzerland (19), Syria (20), Thailand (21), United Kingdom (22), United States (23) and Yugoslavia (24). In our example we chose one of the relations, imports of food and live animals, and two of the attribute variables, secondary school enrollment ratio in 1980 and energy consumption per capita in 1980. In a dichotomous directed graph partitioned into b blocks, we need to estimate b^2 edge probabilities. In our example we will reduce the number of probabilities to $b(b+1)/2$ by symmetrizing the relational variable in the following way: let $x_{ij} = x_{ji} = 1$ if at least one of the following restrictions, $x_{ij} = 1$ or $x_{ji} = 1$, are satisfied. Otherwise $x_{ij} = x_{ji} = 0$.

5.2.1 With two blocks

To ease the subjectivity in our analysis, the following three sets of priors for $\boldsymbol{\eta}$ are entertained expressing various initial beliefs:

1. A uniform distribution over the restricted domain $0 < \eta(1,1) < \eta(0,0) < 1$.

2. $\eta(0,1) \sim \text{beta}(6,3)$, $\eta(0,0) \sim \text{beta}(6,3)$ and $\eta(1,1)$ is uniformly distributed. Repetitive draws of $\eta(0,0)$ and $\eta(1,1)$ are performed until the restriction $0 < \eta(1,1) < \eta(0,0) < 1$ is satisfied.

3. $\eta(0,1) \sim \text{beta }(3,15)$, $\eta(0,0) \sim \text{beta}(6,3)$ and $\eta(1,1)$ is uniformly distributed. Repetitive draws of $\eta(0,0)$ and $\eta(1,1)$ are performed until the restriction $0 < \eta(1,1) < \eta(0,0) < 1$ is satisfied.

Our prime concern is to vary the location of the prior distribution of the edge probabilities between the blocks. Prior 1 represents a prior belief of a person with little knowledge of the trading activity between the blocks. Prior 2 corresponds to a prior belief of a person that assumes a rather high trading activity between the blocks, whereas prior 3 corresponds to a prior belief of a person that assumes a low trading activity between the blocks. Under all three sets of priors $\eta(0,0)$ and $\eta(1,1)$ are dependent, and the vector $(\eta(0,0), \eta(1,1))$ is defined in the region $0 < \eta(1,1) < \eta(0,0) < 1$ independently of $\eta(0,1)$.

The vector of unknown probit parameters is now given by $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$. The hyperparameters in our assumed normally distributed prior is set to $\boldsymbol{\beta}^* = (0,0,0)$ and $\mathbf{B}^* = (\mathbf{z}'\mathbf{z})^{-1}\delta\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and δ is a parameter that reflects our prior uncertainty of $\boldsymbol{\beta}$. As δ increases, the Gibbs sampler will fail to converge; see Figure 6 which shows posterior mean evolution of β_0 for three values of δ. A cause of concern is that an increase of the number of probit model parameters requires an increase of prior information in order to implement this approach. This is due to the nature of the properties of the Markov chain. In the remainder of the analysis in this paper,

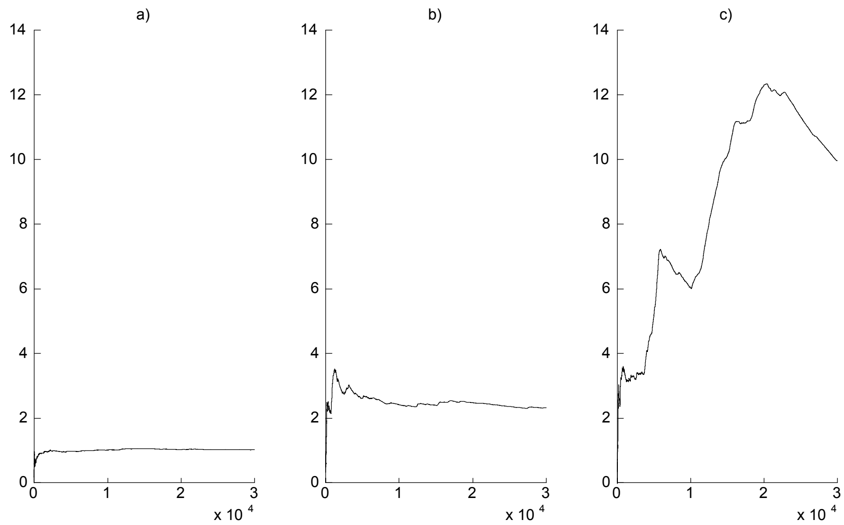

Figure 6: Posterior mean evolution of β_0 for a) $\delta = 10$, b) $\delta = 100$ and c) $\delta = 1000$ based on a Gibbs sampling with $30,000$ iterations.

we assume that our a priori uncertainty regarding $\boldsymbol{\beta}$ corresponds to a value of δ equal to 10.

Note that a standardizing factor $(\mathbf{z}'\mathbf{z})^{-1}$ is included in the covariance matrix to revoke excessive influence of the intercept on the actor block probabilities $\Phi(\mathbf{z}_i'\boldsymbol{\beta})$. Without considering the scale of the predictors we can obtain large positive or negative values of the intercept that will govern the values of $\Phi(\mathbf{z}_i'\boldsymbol{\beta})$ to a large extent. This can, in some extreme cases, lead to the deceptive result that all actors have high posterior probabilities to belong to the same block.

The posterior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are exhibited in Figure 7. It is encouraging to see that the posterior of $\boldsymbol{\beta}$ is insensitive to the choice of the $\boldsymbol{\eta}$-prior. The posteriors of $\boldsymbol{\eta}$ are of course somewhat dependent on the chosen $\boldsymbol{\eta}$-prior as the data set is not large enough to reconciliate the rather disparate prior options.

In Figure 8 the posterior probabilities to belong to block 1 under the model with covariates are plotted against the posterior probabilities to belong to block 1 under the model without covariates. Our prime concern is addressed to the actors located off the diagonal. Since they are mainly located below the diagonal we would expect them to have high covariate values

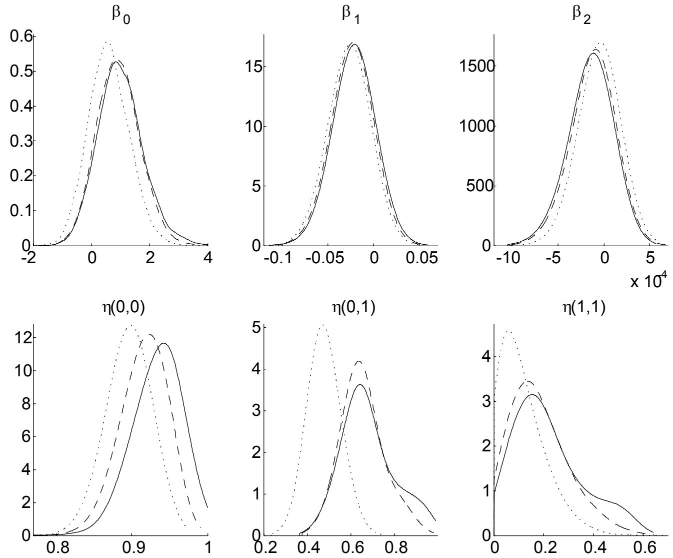

Figure 7: Posterior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ with a normally distributed prior on $\boldsymbol{\beta}$, and prior 1 (solid line), prior 2 (dashed line) and prior 3 (dotted line) on $\boldsymbol{\eta}$.

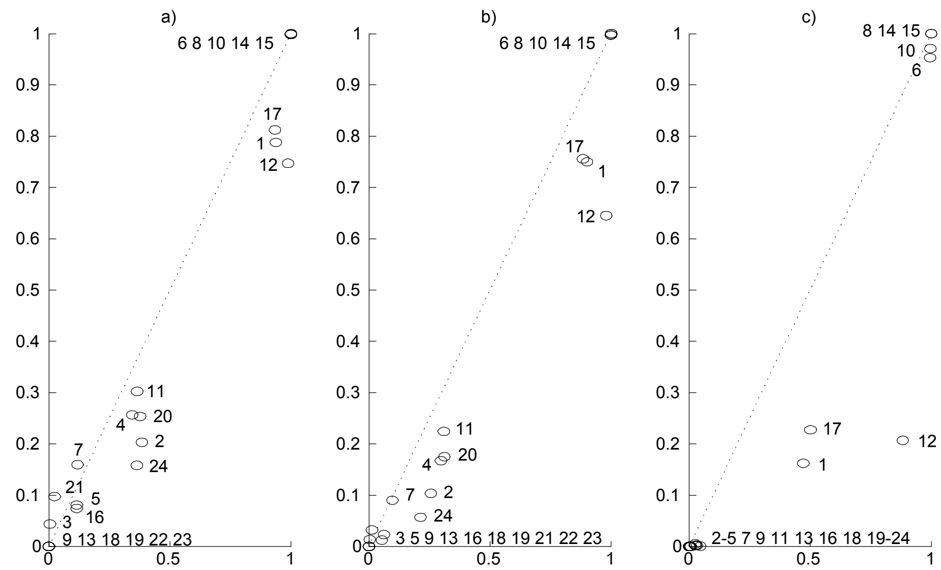

Figure 8: Posterior probabilities to belong to block 1 without covariates (x-axis) and with covariates (y-axis), with our assumed normally distributed prior on β and a) prior 1, b) prior 2 and c) prior 3 on $\boldsymbol{\eta}$.

due to the negativity of the posterior means of β_1 and β_2; see Table 1. This is certainly true for actors 2, 12 and 24, but not for actors 1, 4, 11, 17 and 20 according to Figure 9. This may be explained by the fact that although the posterior means of β_1 and β_2 are negative, their probabilities to attain positive values are quite large according to their posterior distributions in Figure 7. However, by assuming three blocks instead of two, the impact of available covariate information on the posterior block probabilities are more in line with what one would expect.

Parameter	Mean	Std	95% prob. inter.
β_0	1.004	0.791	$(-0.427, 2.719)$
β_1	-0.021	0.019	$(-0.060, 0.016)$
β_2	-1.423$\times 10^{-4}$	2.131$\times 10^{-4}$	$(-6.098 \times 10^{-4}, 2.333 \times 10^{-4})$

Table 1: MCMC results for the parameters $\boldsymbol{\beta}$

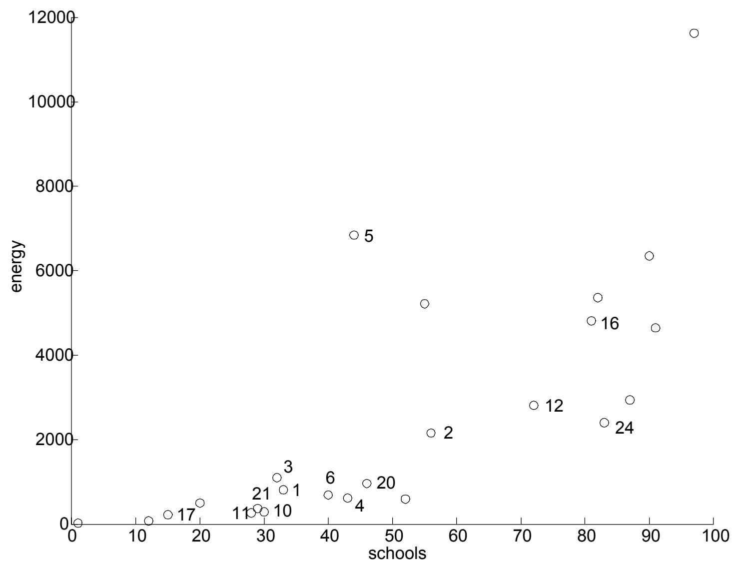

Figure 9: Plot of actor covariates energy against schools.

5.2.2 With three blocks

When the number of blocks is larger than two, besides setting priors on $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, the Bayesian approach involves an additional determination of a set of priors on the covariance matrix Σ. With three blocks, Σ consists of just one unconstrained variance, $\phi + \gamma\gamma'$, and one covariance, γ. Since we in our simulated data set example and real data set example with two blocks, evaluated our proposed model for different prior beliefs of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, we are satisfied in the sequel to present results for various prior information of ϕ and γ. Two sets of priors are considered for ϕ and γ, of which both sets are roughly centered on $\Sigma = \mathbf{I}$. The tightness of the prior is controlled by C and m, small values of C and large values of m give tighter priors; see Equation (6) and Equation (7). Our first prior is non-informative with $C = 100$ and $m = 1$, whereas the second is more informative with $C = 3/4$ and $m = 6$.

With three blocks the array of edge probabilities is given by

$$\boldsymbol{\eta} = \left[\eta\left(0,0\right), \eta\left(1,1\right), \eta\left(2,2\right), \eta\left(0,1\right), \eta\left(0,2\right), \eta\left(1,2\right)\right].$$

We assume a priori that $\eta\left(0,0\right)$, $\eta\left(0,1\right)$ and $\eta\left(0,2\right)$ are beta $(6,3)$ distributed; $\eta\left(1,1\right), \eta\left(2,2\right)$ and $\eta\left(1,2\right)$ are beta $(5,5)$ distributed, beta $(3,6)$

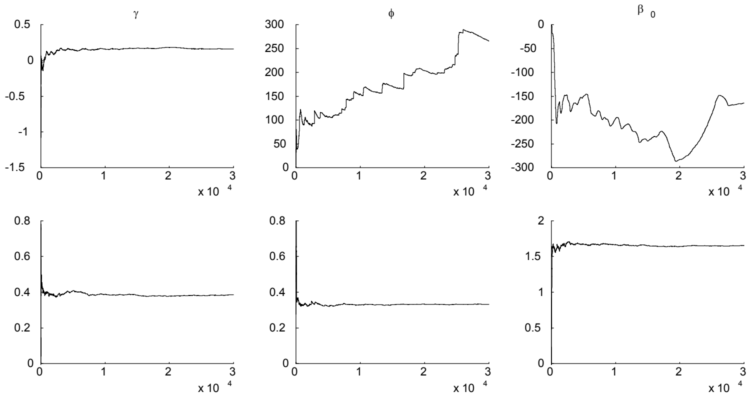

Figure 10: Posterior mean evolution of γ, ϕ and β_0, respectively for two sets of priors of γ and ϕ. In the top row $C = 100$ and $m = 1$, in the bottom row $C = 3/4$ and $m = 6$.

distributed and beta $(2, 8)$ distributed, respectively. Hence, we assume prior to looking at our data a high trading activity within block 0, and between block 0 and the other two blocks. Furthermore, we assume a different degree of low trading activity within block 1, within block 2, and between block 1 and block 2.

Figure 10 exhibits the posterior mean evolution of γ, ϕ and β_0 for the two sets of priors on Σ. Although the Gibbs sequence of γ converges to a value close to 0 under the non-informative set of priors, we see that it is not immediately obvious that the sequences of the other parameters will converge as well. Lack of convergence of ϕ, results in lack of convergence of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ too, here represented by β_0. Under our more informative prior, the Gibbs sampler converges for all the model parameters.

To evaluate the impact on the predictive posterior distributions of block probabilities by including information from actor covariates in the model, we examine Figures 9 and 12. Our immediate concern are those actors whose posterior block probabilities differ between the models, i.e. are located off the diagonal. For example, since β_1 and β_2 are more likely to be negative than positive, see Figure 11 displaying the posterior distributions of $\boldsymbol{\beta}$ and Σ, the probability $\Pr(y_i = 0 \,|\mathbf{z})$ decreases for actors 3 and 21 due to relatively low covariate values and increases for actors 5 and 16 due to relatively

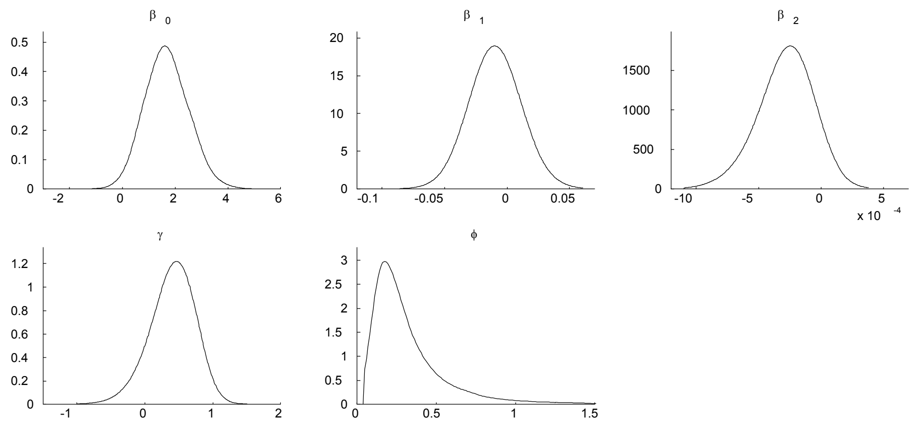

Figure 11: Posterior distributions of $\boldsymbol{\beta}$ and Σ for the informative prior.

high covariate values. Analogously, it follows that $\Pr(y_i = 2\,|\mathbf{z})$ increases for actors 6 and 10 due to low covariate values. The impact of covariates on the predictive posterior probability to belong to block 1, $\Pr(y_i = 1\,|\mathbf{z})$, seems a bit more erratic. Obviously, a decrease in $\Pr(y_i = 1\,|\mathbf{z})$ is explained by an increase in $\Pr(y_i = 0\,|\mathbf{z})$ for actors 5 and 16, whereas it is explained by an increase in $\Pr(y_i = 2\,|\mathbf{z})$ for actors 6 and 10.

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

$$H = -\sum p_y \log p_y$$

where $p_y = \Pr(y)$, and the base of the logarithm is optional. A small value in H implies less uncertainty in the distribution of y. Since the block 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 block seem to increase when covariates are considered. The mean and standard deviation of the difference between H_{nocov} and H_{cov} is 0.182 and 0.164, respectively.

Actor	H_{nocov}	H_{cov}	$H_{\text{nocov}}/H_{\text{cov}}$
1	0.276	0.035	7.89
2	0.410	0.087	4.71
3	0.736	0.323	2.28
4	0.370	0.018	20.56
5	0.537	0.582	0.92
6	0.627	0.450	1.39
7	0.420	0.059	7.12
8	0.268	0.005	53.60
9	0.092	0.033	2.79
10	0.626	0.389	1.61
11	0.348	0.018	19.33
12	0.562	0.643	0.87
13	0.023	0.013	1.77
14	0.243	0.013	18.69
15	0.237	0.001	237.00
16	0.502	0.481	1.04
17	0.296	0.062	4.77
18	0.029	0.036	0.81
19	0.005	0.007	0.71
20	0.323	0.021	15.38
21	0.665	0.167	3.98
22	0.019	0.010	1.90
23	0.008	0.003	2.67
24	0.443	0.231	1.92

Table 2: Entropies of the block affiliation probabilities under both models $\boldsymbol{\beta}$

By examining Figures 8 and 12, we see that actors 6, 8, 10, 14 and 15 are likely to belong to block 1 when the number of blocks are fixed to two and to block 2 when the number of blocks are fixed to three. Actors 9,13, 18, 19, 22, and 23 are likely to belong to block 0 both when the number of blocks are two and three. Actors 1-5, 7, 11, 16, 17, 20, 21 and 24, whose block allocations are not as obvious when the number of blocks are fixed to two, are likely to belong to block 1 when three blocks are considered. Finally, actor 12 has about equal probability to belong to block 0 or block 1 when the number of blocks are two, and block 1 or block 2 when the number of blocks are three. 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 the following six positions (blocks) by using complete link hierarchical clustering:

- $b_0 : 13, 22, 23$
- $b_1 : 4, 5, 11, 18, 24$
- $b_2 : 2, 3, 9, 16, 17, 19, 21$
- $b_3 : 1, 7, 20$
- $b_4 : 6, 10, 12$
- $b_5 : 8, 14, 15$

The constellation of block 0 roughly agrees with b_0, whereas the constellation of block 1 roughly agrees with b_1, b_2 and b_3. Finally, we see by excluding actor 12 from b_4, that the constellation of block 2 equals b_4 and b_5.

We performed our analysis for a predetermined number of blocks, $b = 2$ and $b = 3$. A rather natural extension would be to generalize the analysis to an arbitrary number of blocks. A proper Bayesian analysis would then include the computation of the posterior distribution of the number of blocks.

6 Discussion

An extension to Nowicki and Snijders (2001) Bayesian approach to posterior blockmodeling is presented in this paper. It involves a possibility to use information from covariates on actor level in order to predict the block affiliation of the actors. The block affiliation probabilities are modeled conditional on the covariates via a multinomial probit model. This approach provides computational tractable methods, e.g. Gibbs sampling, to compute posterior distributions of the model parameters and predictive posterior distributions of the block affiliation of each actor, since simulation is only required from standard distributions.

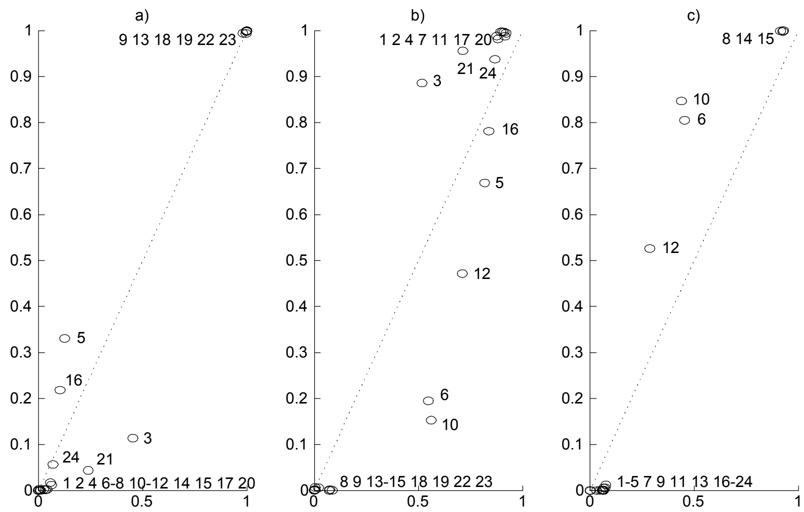

Figure 12: Posterior probabilities to belong to a) block 0; b) block 1; c) block 2, without covariates (x-axis) and with covariates (y-axis), when the hyperparameters are $C = 3/4$ and $\nu = 6$.

Although the predictive posterior distributions of block affiliation of actors is partly governed by prior information on the relational structure, we show in this paper by performing a simulation study, that inclusion of covariate information on actor level in the model can considerably improve the prediction of block affiliation.

The number of parameters increases with the number of covariates and the number of blocks. A cause of concern is requirement of more informative priors on some of the parameters in order to force the Gibbs sampling to converge.

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 block affiliation and the covariate values of the two vertices. By conditioning on the block affiliation and covariate values 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.

Finally, the predictive posterior distributions of block affiliation is computed for a predetermined and fixed number of blocks. As mentioned earlier, an extension would be to generalize the analysis to an arbitrary number of blocks which would include the computation of the posterior distribution of the number of blocks.

References

[1] Aitchison, J. and Bennett, J.A. (1970). Polychotomous quantal response by maximum indicant. *Biometrica.* **57**, 253-262.

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

[3] Fienberg, S.E. and Wasserman, S. (1981). Categorical Data Analysis of Single Sociometric Relations. In *Sociological Methodology-1981,* ed. S. Leinhardt San Fransisco: Jossey-Bass, pp. 156-192.

[4] Frank, O. (1988a). Triad count statistics. *Discrete Mathematics,* ***72***, 141-149.

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

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

[7] Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association,* **81**, 832-842.

[8] 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.

[9] Gilks, W. R., Richardson, S. and 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*, Chapman and Hall, London, pp. 1-19.

[10] Holland, P., Laskey, K.B. and Leinhardt, S. (1983). Stochastic blockmodels: Some first steps. *Social Networks*, **5**, 109-137.

[11] Holland, P. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association, **76**,* 33-65.

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

[13] McCulloch, R., Polson, N.G. and Rossi, P.E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics, **99**,* 173-193.

[14] McCulloch, R. and Rossi, P.E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics,* **64**, 207-240.

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

[16] Richardson, S. and Green, P.J. (1997). On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society,* **59**, 731-792.

[17] Snijders, T.A.B. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent blockstructure. *Journal of classification,* **14**, 75-100.

[18] Wasserman, S. and Anderson, C. (1987). Stochastic *a posteriori* blockmodels: Construction and assessment. *Social Networks*, **9**, 1-36.

[19] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: methods and applications.* New York: Cambridge University Press.

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