Estimation and Prediction for Stochastic Blockstructures

Krzysztof NOWICKI and Tom A. B. SNIJDERS

A statistical approach to a posteriori blockmodeling for digraphs and valued digraphs is proposed. The probability model assumes that the vertices of the digraph are partitioned into several unobserved (latent) classes and that the probability distribution of the relation between two vertices depends only on the classes to which they belong. A Bayesian estimator based on Gibbs sampling is proposed. The basic model is not identified, because class labels are arbitrary. The resulting identifiability problems are solved by restricting inference to the posterior distributions of invariant functions of the parameters and the vertex class membership. In addition, models are considered where class labels are identified by prior distributions for the class membership of some of the vertices. The model is illustrated by an example from the social networks literature (Kapferer's tailor shop).

KEY WORDS: Cluster analysis; Colored graph; Gibbs sampling; Latent class model; Mixture model; Social network.

1. INTRODUCTION AND PREVIEW

Relational structure models are used to describe social, physical, and other phenomena where interactions between units are observed. In the social sciences, these models allow researchers to represent pairwise relational structures of social *actors* (e.g., individuals, organizations, countries), where the relationships are defined by social interactions such as friendship, acquaintance, collaboration, information flow, and combinations of such interactions. A general introduction was given by Wasserman and Faust (1994).

Although the analysis of relational structures has its primary focus on the *pattern of relationships* between the actors involved, the relations often are strongly affected by the *attributes* possessed by the actors (e.g., age, gender, income). Within a group consisting of actors each having the same attributes, the pairwise relationships often exhibit a simpler structure and thus can be modeled using relatively straightforward models. In practical situations, however, we are often interested in studying groups with attributes differing between individual actors, which increases the complexity of the observed pattern of pairwise relationships. The complexity of the situation is further increased by the facts that it often is a priori unclear which attributes influence the relationship patterns, and that these attributes may not have been measured.

In this context of relational data, we assume that a set of n vertices (also referred to as *actors*) is given, and for each ordered pair of two actors i and j (or for a subset of such pairs), a variable y_{ij} can be observed, which we call the relation from i to j. A fruitful approach to modeling relational data has been to suppose that the distribution of the observed relational structure $\mathbf{y} = (y_{ij})_{1 \le i \ne j \le n}$ can be modeled *conditionally on the attribute vector* $\mathbf{x} = (x_1, \ldots, x_n)$, where x_i is the attribute of the ith actor. We allow attributes x_i to take values in a finite set $\mathcal{C}$ and refer to the values as *positions, classes*, or *colors*. More precisely, we assume that the probability distribution of the relation y_{ij} depends on i and j through their attributes x_i and x_j. In the case where the attributes are known and observed, this approach is called a priori blockmodeling, because rearranging the matrix $\mathbf{y}$ according to the values of the attributes divides this matrix into blocks, each consisting of outcomes of an array of identically distributed random variables.

In the deterministic version of this model, the blocks consist entirely of either 0's or 1's. In this situation, Lorrain and White (1971) defined actors with the same attribute values, and thus with the same ties to other actors, to be structurally equivalent. The importance of this equivalence concept was discussed by White, Boorman, and Breiger (1976), and this concept is now established as a major tool in positional role analysis of social networks. Extensive discussions were given by Faust (1988) and Wasserman and Faust (1994).

Fienberg and Wasserman (1981) and Holland, Laskey, and Leinhardt (1983) generalized the deterministic concept of structural equivalence to probabilistic models. They introduced probability models for directed graphs (where y_{ij} can have values 0 and 1) for which the distribution of the relation y_{ij} depends on actor attributes, under the additional assumptions of independence of the *dyads* defined as the pairs (y_{ij}, y_{ji}), and permutation invariance of actors (or vertices) within each class. They called such models *pair-dependent stochastic blockmodels* (see Anderson, Wasserman, and Faust 1992 for a review). In the pair-dependent blockmodel, the actors belonging to the same class are called *stochastically equivalent* in the sense that the probabilities of the relationships with all other actors are the same for all actors in the same class.

Incorporation of stochastic equivalence in a model for relational data is much more difficult when attributes cannot be observed and the class structure can be identified only a posteriori based on the observed relational data $\mathbf{y}$. Wasserman and Anderson (1987) proposed an a posteriori blocking procedure in the framework of the p_1 family (Holland and Leinhardt 1981). Snijders and Nowicki (1997) studied an a posteriori blockmodel for undirected graphs assuming that $|\mathcal{C}| = 2$ and that the probability of an edge between two actors depends only on the classes to which the actors belong. For small graphs, they proposed maximum likelihood (ML) estimation of the parameters based on the EM algorithm, and for large

Krzysztof Nowicki is Associate Professor, Department of Statistics, University of Lund, S-220 07 Lund, Sweden (E-mail: *krzysztof.nowicki@stat.lu.se*). Tom Snijders is at ICS/Department of Statistics and Measurement Theory, University of Groningen, 9712 TS Groningen, The Netherlands, (E-mail: *t.a.b.snijders@ppsw.rug.nl*). Nowicki's research was supported by the Swedish Council for Research in the Humanities and Social Sciences.

graphs, they proposed a Bayesian estimation method based on Gibbs sampling.

In this article we extend the approach of Snijders and Nowicki (1997) to the case where relations can be directed and can have an arbitrary set of possible values, and where the number of classes is arbitrary. The vector of attributes $\mathbf{x}$ specifying the class structure is considered to be *unobserved* (*latent*). Conditional on the vector $\mathbf{x}$, we model $\mathbf{y}$ using a generalization of the pair-dependent stochastic blockmodel. The model can be regarded as a mixture model because the classes to which the vertices belong are assumed to be outcomes of random variables X_i.

Because it is essential to avoid making restrictive assumptions about the dependence or independence between the reciprocal relations Y_{ij} and Y_{ji}, our unit of analysis is the dyad (Y_{ij}, Y_{ji}). We call the set of possible values for (Y_{ij}, Y_{ji}) the *alphabet* $\mathcal{A}$ of pairwise relations. Because the alphabet is arbitrary, the model allows as data not only graphs and digraphs, but also tournaments, signed graphs, valued digraphs, and so on. The method proposed also allows missing values and structurally predefined relations ("structural zeros").

The parameter estimation method is based on a Bayesian approach. The prior distribution for parameters is taken to be a product of independent Dirichlet distributions for the color distribution and for the relationships between and within classes of the given colors. The posterior distribution is estimated using the Gibbs sampler. (See Gelman, Carlin, Stern, and Rubin 1995 for Gibbs sampling in the context of other mixture models). The analysis is directed especially at the posterior predictive distribution of the vector of vertex colors, $\mathbf{X}$.

As in all mixture models, we have the problem of *nonidentifiability* of parameters, because the color labels $1, \ldots, c$ are not identified; only the partition defined by the coloring is identified. Rather than using, for example, an order restriction on the parameters, we address this problem by considering only the posterior distributions of those functions of parameters and the latent attributes that are invariant with respect to relabeling the classes.

Finally, we also treat the situation where prior information is available, represented by noninvariant prior distributions for the colors or for the relations. Such priors can be used to identify the class labels, thereby circumventing the nonidentifiability problems. Two situations are treated: prior knowledge about vertex colors and prior knowledge about the block structure; that is, the probability distributions of the relationships between vertices of the various colors. An example of the latter is a center–periphery model, which has two classes: the center, with many within-class relations, and the periphery, with low within-class relations and a moderate number of relations with the center.

2. DISCRETE RELATIONAL DATA

In this section we describe the data structures studied. Basic is the set of n vertices labeled $1, \ldots, n$. Throughout this article we assume that identities of vertices are distinguishable, and take for granted the term "labeled" for graphs, digraphs, and so on. A relational structure is assumed on this set of vertices. (Loops, or relations of a vertex with itself, are not considered.) Examples are graphs and digraphs, but we consider the more general case where the relation from vertex i to vertex j takes a value from a finite set $\boldsymbol{\alpha} = \{a_1, a_2, \ldots, a_R\}$. For instance, data consisting of two digraphs on the same set of vertices can be represented as a directed relational structure where $\boldsymbol{\alpha}$ has $R = 4$ elements.

To allow missing data or structurally predefined relations (i.e., the possibility that there are pairs of vertices between which the relation is not observed or which are for other reasons excepted from the model), we let

$$\mathcal{N} \subset \mathcal{N}_0 = \{(i, j) \in \{1, \ldots, n\}^2 \mid i \neq j\} \tag{1}$$

denote the set of pairs for which the relation is observed and to which the probability model applies. We assume that $\mathcal{N}$ is *symmetric* in the sense that if $(i, j) \in \mathcal{N}$, then also $(j, i) \in \mathcal{N}$. If a structure has no missing data, then $\mathcal{N} = \mathcal{N}_0$.

The set $\boldsymbol{\alpha} = \{a_1, a_2, \ldots, a_R\}$ is the set of values of the relation from one vertex to another. One of the aims of our analysis is to explicitly take into account the mutual dependence of the relation from i to j and the relation in the reverse direction, from j to i. This pair of relations takes values in the set $\boldsymbol{\alpha}^2 = \{\mathbf{a} = (a_t, a_v) \mid a_t, a_v \in \boldsymbol{\alpha}\}$.

It is not necessary that all elements of $\boldsymbol{\alpha}^2$ can occur as pairwise relations. The subset $\mathcal{A}$ of $\boldsymbol{\alpha}^2$ containing the potential values of pairwise relations is called the *alphabet* of pairwise relations. Thus, for every pair $(i, j) \in \mathcal{N}$, there is an $\mathbf{a} = (a_t, a_v) \in \mathcal{A}$ such that the relation from i to j is a_t and the relation from j to i is a_v. We then write $y_{ij} = \mathbf{a}$. In a slight deviation from the notation suggested by Section 1, this y_{ij} represents the relation from i to j as well as the relation from j to i. The assumption about the symmetry of $\mathcal{N}$ implies that for each $(a_t, a_v) \in \mathcal{A}$, it holds that $(a_v, a_t) \in \mathcal{A}$.

The following list of alphabets $\mathcal{A}$ gives examples of various well-known relational structures:

- For a graph, we have $\boldsymbol{\alpha} = \{0, 1\}$ and $\mathcal{A} = \{(0, 0), (1, 1)\}$.
- For a digraph, we have $\boldsymbol{\alpha} = \{0, 1\}$ and $\mathcal{A} = \{0, 1\}^2$.
- For a signed digraph, we have $\boldsymbol{\alpha} = \{0, -, +\}$ and $\mathcal{A} = \{0, -, +\}^2$.
- For a tournament, we have $\boldsymbol{\alpha} = \{0, 1\}$ and $\mathcal{A} = \{(0, 1), (1, 0)\}$.
- A tournament where ties are allowed can be represented by $\boldsymbol{\alpha} = \{0, -, +\}$ and $\mathcal{A} = \{(-, +), (+, -), (0, 0)\}$.

An important operator on $\mathcal{A}$ is the reflection operator π defined by

$$\pi(a_t, a_v) = (a_v, a_t).$$

This operator is a square root of the identity $\pi^2(\mathbf{a}) \equiv \mathbf{a}$. The definition of the alphabet $\mathcal{A}$ implies that it is closed under reflection $\pi(\mathcal{A}) = \mathcal{A}$.

The alphabet $\mathcal{A}$ can be partitioned into two subsets, $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$, where $\mathcal{A}_0$ is the set of *symmetric* (or *even*) *relations* defined by

$$\mathcal{A}_0 = \{\mathbf{a} \in \mathcal{A} \mid \pi(\mathbf{a}) = \mathbf{a}\},$$

whereas $\mathcal{A}_1$ is the set of *asymmetric* (or *odd*) *relations*

$$\mathcal{A}_1 = \mathcal{A} \backslash \mathcal{A}_0 = \{\mathbf{a} \in \mathcal{A} \mid \pi(\mathbf{a}) \neq \mathbf{a}\}.$$

We do not need the entire set $\mathcal{A}_1$ to describe asymmetric relations when vertices are identified only by their colors. For example, for arbitrary $\mathbf{a} \in \mathcal{A}_1$, event $E_1 \equiv \{\mathbf{a}$ occurs from some vertex of color k to some vertex of color $h\}$ is equivalent to event $E_2 \equiv \{\pi(\mathbf{a})$ occurs from some vertex of color h to some vertex of color $k\}$. Thus we find it necessary, in connection with the introduction of non-redundant parameterization (see Sec. 3), to partition set $\mathcal{A}_1$ into two disjoint half sets such that for each $\mathbf{a} \in \mathcal{A}_1$, one half set contains $\mathbf{a}$ and the other contains $\pi(\mathbf{a})$. (There are many ways to choose this partition in two half sets, and any will do.) Denoting these half sets by $\mathcal{A}_{10}$ and $\mathcal{A}_{11}$, we have

$$\mathcal{A}_1 = \mathcal{A}_{10} \cup \mathcal{A}_{11}, \mathcal{A}_{10} \cap \mathcal{A}_{11} = \varnothing, \pi(\mathcal{A}_{10}) = \mathcal{A}_{11},$$

and we define

$$\mathcal{A}' = \mathcal{A}_0 \cup \mathcal{A}_{10}.$$

We denote $r = \sharp\mathcal{A}$, $r_0 = \sharp\mathcal{A}_0$, and $r_1 = \frac{1}{2}\sharp\mathcal{A}_1 = \sharp\mathcal{A}_{10} = \sharp\mathcal{A}_{11}$.

A relational structure on the set $\mathcal{N}$ can be represented by its (*generalized*) *adjacency matrix* $\mathbf{y} = (y_{ij})_{(i,j)\in\mathcal{N}}$. Recall that, in deviation from usual practice, element y_{ij} of the matrix defines the relation from i to j as well as that from j to i. Clearly,

$$y_{ji} = \pi(y_{ij}). \tag{2}$$

The second aspect of the structure is a discrete vertex characteristic. We assume that the set of vertices $\{1, \ldots, n\}$ is partitioned into c categories labeled $1, \ldots, c$. The category of vertex i, also called its *class* or *color*, is denoted by x_i. The set of colors is denoted by $\mathcal{C} = \{1, \ldots, c\}$. The colors are collected in the vector $\mathbf{x} = (x_i)_{i=1}^n$.

Summarizing, in this article we consider a set $\mathcal{N}$ of ordered pairs of vertices between which relations are given, with the vertices belonging to c categories or colors and with a relational structure where the pair of relations from i to j and from j to i takes values in a finite alphabet $\mathcal{A}$. This is called a *colored relational structure* and can be represented by the array $(\mathbf{y}, \mathbf{x})$. We consider random structures where the set of pairs of vertices $\mathcal{N}$, the set of colors $\mathcal{C}$, and the alphabet $\mathcal{A}$ are fixed and the structure $(\mathbf{y}, \mathbf{x})$ is stochastic. Further, $\mathbf{y}$ is observed but $\mathbf{x}$ is not. We denote random variables by capital letters; hence the random adjacency matrix and the random color vector are denoted by $\mathbf{Y}$ and $\mathbf{X}$.

3. STOCHASTIC STRUCTURES WITH RANDOMLY COLORED VERTICES

To define stochastic blockmodels for the colored relational structure defined above, we take an approach similar to mixture modeling, and assume that the random colors X_i are iid random variables with probability

$$\Pr(X_i = k) = \theta_k \tag{3}$$

for color $k \in \mathcal{C}$. Hence the joint distribution of $\mathbf{X}$ is defined by

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \theta_1^{m_1} \ldots \theta_c^{m_c},$$

where

$$m_k = \sum_{i=1}^{n} I(x_i = k) \tag{4}$$

denotes the number of vertices with color k.

The observations consist of the relations between the vertices. The model for the relationships between the vertices depends on vertex coloring in the following way. Given the vector of vertex colors $\mathbf{X} = \mathbf{x}$, the random vectors $\mathbf{Y}_{ij}$ for $(i, j) \in \mathcal{N}$ with $i < j$ are independent, and the probabilities are

$$\Pr(\mathbf{Y}_{ij} = \mathbf{a} \mid \mathbf{X} = \mathbf{x}) = \eta_{\mathbf{a}}(x_i, x_j), \tag{5}$$

where the array

$$\boldsymbol{\eta} = \eta_{\mathbf{a}}(k, h) \quad \text{for} \quad \mathbf{a} \in \mathcal{A}, \qquad k, h \in \mathcal{C}$$

of color-dependent dyad probabilities satisfies

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

Corresponding to the symmetry in the relations expressed by (2), the probabilities $\boldsymbol{\eta}$ must be invariant with respect to reflection,

$$\eta_{\mathbf{a}}(k, h) = \eta_{\pi(\mathbf{a})}(h, k). \tag{6}$$

This implies a redundancy in parameters $\eta_{\mathbf{a}}(k, h)$. A nonredundant parameterization is obtained by defining the parameter $\boldsymbol{\eta}$ as comprising $\eta_{\mathbf{a}}(k, k)$ for $\mathbf{a} \in \mathcal{A}', k \in \mathcal{C}$ (representing the probability distributions of relations between vertices of the same color) and $\eta_{\mathbf{a}}(k, h)$ for $\mathbf{a} \in \mathcal{A}, k < h$ (probability distributions of relations between vertices of different colors). The conditional distribution of relations $\mathbf{Y}$ given the vector of colors $\mathbf{x}$ is given in the following form (using only the nonredundant parameters contained in $\boldsymbol{\eta}$):

$$\Pr(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \left(\prod_{\mathbf{a}\in\mathcal{A}} \prod_{1\le k<h\le c} (\eta_{\mathbf{a}}(k, h))^{e_{\mathbf{a}}(k,h)} \right) \times \left(\prod_{\mathbf{a}\in\mathcal{A}'} \prod_{k=1}^{c} (\eta_{\mathbf{a}}(k, k))^{e_{\mathbf{a}}(k,k)} \right), \tag{7}$$

where $e_{\mathbf{a}}(k, h), \mathbf{a} \in \mathcal{A}, 1 \le k \le h \le c$, are the relation counts for block (k, h), defined by

$$e_{\mathbf{a}}(k, h) = (1 + I\{k = h\} I\{\mathbf{a} \in \mathcal{A}_0\})^{-1} \times \sum_{(i,j)\in\mathcal{N}} I\{y_{ij} = \mathbf{a}\} I\{x_i = k\} I\{x_j = h\}, \tag{8}$$

and $I\{A\}$ is defined as 1 if condition A is satisfied, and 0 otherwise.

This gives an extension of Holland et al's (1983) concept of a *pair-dependent stochastic blockmodel*; they restricted their study to digraphs; that is, $\boldsymbol{\alpha} = \{0, 1\}$ and $\mathcal{A} = \{0, 1\}^2$.

The stochastic blockmodel is then given by the joint distribution of $(\mathbf{Y}, \mathbf{X})$,

$$\Pr(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = \theta_1^{m_1} \cdots \theta_c^{m_c} \times \left(\prod_{\mathbf{a} \in \mathcal{A}} \prod_{1 \le k < h \le c} (\eta_{\mathbf{a}}(k,h))^{e_{\mathbf{a}}(k,h)} \right) \times \left(\prod_{\mathbf{a} \in \mathcal{A}'} \prod_{k=1}^{c} (\eta_{\mathbf{a}}(k,k))^{e_{\mathbf{a}}(k,k)} \right). \quad (9)$$

Despite its simplicity, this model provides a powerful tool in modeling relational structures. It allows one to represent the effect of unobserved heterogeneity of individual positions or preferences on the pattern of pairwise relations. The fact that the heterogeneity is modeled by stochastic membership of the c classes makes it, in terms of cluster analysis, analogous to a mixture (or random partition) model rather than a discrete classification (or fixed partition) model (cf. Bock 1996a, b). Fixed partition models contain the n class membership indicators of vertices as incidental parameters, which may lead to inconsistent estimation. In our model, on the other hand, the number of parameters is fixed at $(1/2)c^2(r-1) + (1/2)c(r_0+1) - 1$, and the colors x_i occur as outcomes of random variables.

Various properties of the stochastic blockmodel have been studied in the literature. Without attempting to give a complete review, we mention the following. Frank and Harary (1982), motivated by entropy calculations, considered the case of an undirected colored graph and discussed statistical inference for $s_2 = \sum_{k=1}^{c} \theta_k^2$ and $s_3 = \sum_{k=1}^{c} \theta_k^3$ under the assumption that $\eta_{kl} = (1-\alpha)\delta_{kl} + \beta(1-\delta_{kl})$, where η_{kl} denotes the probability of an edge between vertices of colors k and h. They proposed several moment-based estimators for s_2 and s_3 under various restrictions on $\theta_1, \ldots, \theta_c, \alpha$, and β. Frank (1988a) obtained the expectation and the variance for the number of edges and for the vector of triad counts in the undirected graph. Janson and Nowicki (1991) studied the asymptotic distributions of the vector of suitable normalized subgraph counts for undirected and directed colored graphs and random colored tournaments. Frank (1988b) and Wellman, Frank, Espinoza, Lundquist, and Wilson (1991) considered statistical inference for certain models for randomly colored graphs assuming that the edges as well as the colors are observed.

In this article we consider a structure $(\mathbf{Y}, \mathbf{X})$ with randomly colored vertices as given by (3), assuming that only the relational structure $\mathbf{Y}$ can be observed; that is, the color vector $\mathbf{X}$ is unobserved (latent). Thus the probability of observing edge pattern $\mathbf{y}$ can be written as

$$\Pr(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{\mathbf{x} \in \mathcal{C}^n} \Pr(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\eta}). \quad (10)$$

Given this a posteriori blockmodel, we wish to estimate the parameters $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ and predict the unobserved coloring $\mathbf{x}$.

Wasserman and Anderson (1987) pioneered statistical posterior blockmodeling by considering it for digraphs in the context of Holland and Leinhardt's (1981) p_1 model. The p_1 family of distributions for digraphs includes two parameters for each vertex, called the productivity parameter (related to the number of outgoing relations) and the popularity parameter (related to the number of incoming relations), as well as the reciprocity parameter. This implies certain expressions for the elements of $\boldsymbol{\eta}$ as functions of $2c$ parameters; for instance, for the case of digraphs, this number can be compared to $\boldsymbol{\eta}$ in our model having $3c(c+1)/2 - 1$ parameters.

The procedure proposed by Wasserman and Anderson proceeds by first fitting the p_1 model to digraph data, then grouping together those vertices that have approximately the same ML estimates of the productivity and popularity parameters, and finally fitting the pair-dependent stochastic blockmodel.

Basing posterior blockmodeling on the p_1 model has a disadvantage resulting from its restrictive nature. In the p_1 model, vertices with a high productivity parameter have relatively high probabilities of outgoing ties, and vertices with a high popularity parameter have relatively high probabilities of incoming ties, for all other vertices. This excludes, for example, the important case of digraphs with classes of vertices, where the density of relations is high within classes and low between classes. When it is not known a priori that the special structure of the p_1 model holds, it is safer to use our more general model, which does not make this specific assumption.

Snijders and Nowicki (1997) considered posterior blockmodels for undirected graphs; that is, $\mathcal{A} = \{(0,0), (1,1)\}$ with $c = 2$ latent classes. They elaborated various statistical estimation and prediction procedures, including the EM algorithm, the profile likelihood, the conditional predictive likelihood, and Gibbs sampling. In addition, they proved that asymptotically it is possible, under certain weak conditions, to recover the unobserved vertex colors correctly with probability tending to 1. In their study it appeared that for graphs with about 15 or more vertices, a Bayesian approach with Gibbs sampling was the only feasible one among the methods considered.

This article generalizes the Bayesian approach of Snijders and Nowicki (1997) to the analysis of the a posteriori block model defined by (9) and (10). A prior density function $f(\boldsymbol{\theta}, \boldsymbol{\eta})$ is assumed for the parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$. The recovery of the block structure is based on the *posterior predictive distribution*, which is the conditional distribution of $\mathbf{X}$ given the data $\mathbf{y}$,

$$\Pr(\mathbf{x} \mid \mathbf{y}) = \int f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} \mid \mathbf{y}) d\boldsymbol{\eta} d\boldsymbol{\theta}, \quad (11)$$

and inference about the parameters is based on the *posterior distribution*,

$$f(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \mathbf{y}) = \sum_{\mathbf{x}} f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} \mid \mathbf{y}). \quad (12)$$

Gibbs sampling is used to obtain the conditional distribution $f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x} \mid \mathbf{y})$.

4. IDENTIFIABILITY PROBLEMS AND INVARIANT PARAMETERS

The parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$ in the joint distribution of parameters and data, given by the probability density function $f(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{x}, \mathbf{y})$, are not identifiable, because there are several distinct parameter vectors associated with the same probability distribution for $\mathbf{Y}$. This is because what matters is the partition defined by $\mathbf{x}$, not the color labels $1, \ldots, c$.

To describe this more formally, we define the transformations under which the distribution of $\mathbf{Y}$ is invariant. Let $\mathcal{S}$ denote the group of permutations of $\{1, \ldots, c\}$, and for $s \in \mathcal{S}$ define

$$h_s(\boldsymbol{\theta}, \boldsymbol{\eta}) = \big((\theta_{s(i)})_{i=1}^{c}, (\eta_{\mathbf{a}}(s(k), s(k)))_{\mathbf{a}\in\mathcal{A}', 1\le k\le c}, \times (\eta_{\mathbf{a}}(s(k), s(h)))_{\mathbf{a}\in\mathcal{A}, 1\le k<h\le c}\big); \quad (13)$$

that is, h_s is the function operating on the parameter vector $(\boldsymbol{\theta}, \boldsymbol{\eta})$ in such a way that $\boldsymbol{\theta}$ is permuted by s while the elements of $\boldsymbol{\eta}$ are permuted correspondingly. Then

$$\Pr(\mathbf{Y} = \mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = \Pr(\mathbf{Y} = \mathbf{y} \mid h_s(\boldsymbol{\theta}, \boldsymbol{\eta})) \text{ for all } s \in \mathcal{S}.$$

The invariance of the distribution of $\mathbf{Y}$ affects the statistical analysis concerning the latent coloring of the vertices. Define

$$s(\mathbf{x}) = \big(s(x_1), \ldots, s(x_n)\big), \quad (14)$$

the new coloring of the vertices induced by s. This transformation preserves the partition of the vertices. The definition of h_s implies that

$$\Pr(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta}) = \Pr\big(s(\mathbf{x}) \mid \mathbf{y}, h_s(\boldsymbol{\theta}, \boldsymbol{\eta})\big).$$

Hence, on the basis of observing only $\mathbf{y}$, it is impossible to distinguish between the colorings given by $\mathbf{x}$ and $s(\mathbf{x})$. The colors may be distinguishable, but their labels are not identifiable.

Because the statistical model is invariant under relabeling of the colors, when the prior distribution is invariant, then the posterior also is invariant. Consequently, the posterior predictive probability $\Pr(X_i = k \mid \mathbf{y})$ is independent of k, and thus must be equal to $1/c$. Thus the marginal conditional distribution of X_i given $\mathbf{y}$ tells us nothing, and we need either more detail about the joint distribution of $\mathbf{X}$ or a noninvariant prior.

The aforementioned problems constitute a trivial kind of nonidentifiability, which occurs in all finite mixture and latent class models (see, e.g., Gelman et al. 1995, p. 422). It does not present a problem in ML estimation, because an arbitrary labeling of the classes may be chosen. Now we describe how to deal with this nonidentifiability in a Bayesian approach.

One way to deal with this invariance is to put restrictions on the parameters (or a subset of the parameters). This was proposed by, for example, Gelman et al. (1995, p. 422). One possible restriction is $\theta_1 < \theta_2 < \cdots < \theta_c$, provided that one does not object to excluding distributions for which some of the classes k have equal values of θ_k. This restriction does not work well from a statistical standpoint if some of the color probabilities θ_k are so similar that they cannot be distinguished well empirically (to be understood, e.g., in the sense that their posterior confidence intervals overlap). In such a case the use of this order restriction leads to an imperfect identification of the k classes and, consequently, potentially incorrectly interpreted results. Another possibility is an order restriction on some subvector of the $\boldsymbol{\eta}$ parameter to achieve identifiability. For example, for an undirected graph, where $\mathcal{A} = \{(0,0), (1,1)\}$, one could obtain an identifiable model by assuming that $\eta_{(1,1)}(k,k)$ is an increasing value of k. Such a restriction can fail for similar reasons. A general problem with identifying parameters by means of such order restrictions is that, because of the nature of the relational data, there are parameter values that cannot be made identifiable by any order restriction. This is shown by the following example.

Example. Consider an undirected graph (i.e., $\mathcal{A} = \{(0,0), (1,1)\}$) and denote $\eta(k,h) = \eta_{(1,1)}(k,h)$. Suppose that there are $c = 2$ classes of equal probability $\theta_1 = \theta_2 = .5$, whereas the block structure is defined by the relation (1,1) having a high probability within classes and a low probability between classes. Specifically, one could consider $\eta(1,1) = \eta(2,2) = .8$ and $\eta(1,2) = \eta(2,1) = .3$. If n is not too small, then such classes are very well distinguishable empirically, but they are similar in all respects and cannot be distinguished by an order restriction on the parameters.

To generalize this example somewhat, note that a model for an undirected graph with two classes always has $\eta(1,2) = \eta(2,1)$ [see (6)]. An order restriction on the parameters will lead to practical problems whenever both the posterior distributions of θ_1 and θ_2, and also the posterior distributions of the within-class relation probabilities $\eta(1,1)$ and $\eta(2,2)$, are poorly separated.

As a way of dealing with the problem of unidentifiability of class labels in the case where there is no identifying prior information, we propose to consider only the posterior distributions of functions of $(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{X})$ that are invariant with respect to relabeling the classes; that is, functions that are invariant with respect to the transformations $h_s(\boldsymbol{\theta}, \boldsymbol{\eta})$ and $s(\mathbf{X})$ defined by (13) and (14). Important instances of such invariant functions include the following:

- The indicator function

$$I\{X_i = X_j\} \quad (15)$$

  that two given vertices i and j have the same color
- The probability of the class to which a given vertex i belongs (i.e., θ_{X_i}), which can be defined more elaborately as

$$\theta_{X_i} = \sum_{k=1}^{c} \theta_k I\{X_i = k\}. \quad (16)$$

  More concisely, this could be called the class probability of vertex i (If vertex i tends to be in a class with many other vertices, then θ_{X_i} will tend to be high, irrespective of the label of this class.)
- The probability distribution of the relation Y_{ij} between vertices i and j, given the color vector $\mathbf{X}$ and the probability vector $\boldsymbol{\eta}$, denoted by $\eta_{\mathbf{a}}(X_i, X_j)$ [cf. (5)]:

$$\eta_{\mathbf{a}}(X_i, X_j) = \Pr(Y_{ij} = \mathbf{a} \mid \mathbf{X}, \boldsymbol{\eta}) = \sum_{1\le k,\, h\le c} \eta_{\mathbf{a}}(k,h) I\{X_i = k, X_j = h\}. \quad (17)$$

The posterior means of (15) can be arranged in the matrix of the posterior predictive probabilities,

$$\big(\Pr(X_i = X_j \mid \mathbf{y})\big)_{1\le i\ne j\le n}. \quad (18)$$

This matrix will show the extent to which the data imply a clear block structure; if there is a clear block structure, then

some elements of (18) will be close to 1, with the others close to 0. This is reflected in a low value of H_x, defined in (25).

If, as we propose later, an invariant Dirichlet prior is used for $\boldsymbol{\theta}$ with large values for the parameters T_k, then the posterior means of (16) are strongly drawn toward the prior means $1/c$. These are then of less practical interest.

The posterior means of (17) can be arranged in the three-way array

$$\left(\mathrm{E}(\eta_{\mathbf{a}}(X_i, X_j) \mid \mathbf{y})\right)_{1\le i\ne j\le n,\, \mathbf{a}\in\mathcal{A}'}. \tag{19}$$

This array exhibits the block structure in terms of the posterior probabilities of the various relations $\mathbf{a} \in \mathcal{A}'$.

5. GIBBS SAMPLING TO OBTAIN THE POSTERIOR DISTRIBUTION

Gibbs sampling (explained by, e.g., Gelman et al. 1995 and Gilks, Richardson, and Spiegelhalter 1996) is a simulation method that can be used to approximate the posterior distribution to any desired precision. The Gibbs sampler is an iterative simulation scheme that operates by repeatedly drawing in turn each of a set of unknown random variables or vectors, each conditionally on the values of all of the other random variables. Our application of Gibbs sampling is close to applications for latent class and finite mixture models (e.g., Gelman et al. 1995, chap. 16; Robert 1996).

We apply this scheme to $(\boldsymbol{\theta}, \boldsymbol{\eta}), X_1, \ldots, X_n$, treating $(\boldsymbol{\theta}, \boldsymbol{\eta})$ as a single random vector with the prior density $f(\boldsymbol{\theta}, \boldsymbol{\eta})$. Given current values $\mathbf{X}^{(p)}, \boldsymbol{\theta}^{(p)}$, and $\boldsymbol{\eta}^{(p)}$, the next values $\mathbf{X}^{(p+1)}, \boldsymbol{\theta}^{(p+1)}$, and $\boldsymbol{\eta}^{(p+1)}$ are determined as follows:

1. $\boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)}$ is drawn from the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\eta})$, given the complete data $(\mathbf{X}^{(p)}, \mathbf{y})$.
2. For each value $i = 1, \ldots, n$ in turn, $X_i^{(p+1)}$ is drawn from the conditional distribution of X_i given the values $\boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)}, \mathbf{y}, X_h^{(p+1)}$ for $h = 1, \ldots, i-1$, and $X_h^{(p)}$ for $h = i+1, \ldots, n$.

Because step 1 does not depend on the current value of $(\boldsymbol{\theta}, \boldsymbol{\eta})$, only the initial value $\mathbf{X}^{(1)}$ is important as a starting condition.

It follows from the general theory of Gibbs sampling (e.g., Gelman et al. 1995; Gilks et al. 1996) that for this iteration scheme, irrespective of the starting values, the distribution of $(\boldsymbol{\theta}^{(p)}, \boldsymbol{\eta}^{(p)})$ converges to the posterior distribution, given the observed data $\mathbf{y}$, and the distribution of $\mathbf{X}^{(p)}$ converges to the posterior predictive distribution.

The Gibbs sampler is so convenient here thanks to the simplicity of the conditional distributions from which $\mathbf{X}^{(p+1)}$ and $(\boldsymbol{\theta}^{(p+1)}, \boldsymbol{\eta}^{(p+1)})$ are drawn. The conditional distribution of X_i given $\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{y}$, and $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, needed in step 2, can be derived from (9). For $i \in \{1, \ldots, n\}, \mathbf{a} \in \mathcal{A}, k \in \mathcal{C}$, define

$$d_{\mathbf{a}}(i,k) = \sum_{j:(i,j)\in\mathcal{N}} I\{y_{ij} = \mathbf{a}\} I\{x_j = k\}. \tag{20}$$

It can be concluded from (8) that the conditional probability distribution of X_i is determined by

$$\Pr(X_i = k \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta}, \{X_j\}_{j\ne i}) = Q\theta_k \prod_{\mathbf{a}\in\mathcal{A}} \prod_{h=1}^{c} (\eta_{\mathbf{a}}(k,h))^{d_{\mathbf{a}}(i,h)}, \tag{21}$$

where Q is a constant not depending on k. This equation determines the probability distribution from which $X_i^{(p+1)}$ is to be drawn in step 2 of the Gibbs sampler.

For step 1, we need the specification of the prior distribution, $f(\boldsymbol{\theta}, \boldsymbol{\eta})$. If the number of vertices is large enough and the prior relatively flat, then the prior distribution does not have a large influence on the results of the statistical analysis. In most cases it will be reasonable to assume prior independence between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. If an invariant and flat prior is required to indicate absence of special prior information, then a natural choice is the uniform distribution on the parameter space. For other priors, a large and convenient class is formed by the Dirichlet distributions (multivariate beta distributions), which include the uniform distribution as a special case.

Note that the Dirichlet prior $\mathrm{D}(T, \ldots, T)$ for $\boldsymbol{\theta}$ has a stronger tendency to favor unequal class sizes m_k when T is smaller, and that the uniform distribution corresponds to $T = 1$. For low values of T there is the risk that in early stages of the iteration process, the parameters will be trapped in a region where some of the class sizes are almost 0 and the corresponding elements of $(\boldsymbol{\theta}, \boldsymbol{\eta})$ have a distribution close to uniform. Therefore, in practice the uniform prior distribution for $\boldsymbol{\theta}$ represents a model with a nonnegligible probability of having less than c classes. A Dirichlet prior $\mathrm{D}(T, \ldots, T)$ with a higher value for T is a more adequate representation of the prior idea that there are c classes. We have had good experience using $T = 100c$. Note that such a prior distribution is still permutation invariant, and that it will allow small class sizes if the observations point that way.

Some care must be taken in specifying the prior distribution for $\boldsymbol{\eta}$ because of the redundancy (6) in the η parameters. For $k < h$, $\boldsymbol{\eta}(k,h) = (\eta_{\mathbf{a}}(k,h))_{\mathbf{a}\in\mathcal{A}}$ is an unconstrained r-dimensional probability vector. For $k = h$, this vector is subject to the constraint that $\eta_{\mathbf{a}}(k,k) = \eta_{\pi(\mathbf{a})}(k,k)$. This constraint amounts to r_1 equality relations between elements of $\boldsymbol{\eta}(k,k)$. The definition

$$\eta_{\mathbf{a}}^{(0)}(k,k) = \begin{cases} \eta_{\mathbf{a}}(k,k) & (\mathbf{a}\in\mathcal{A}_0) \\ 2\eta_{\mathbf{a}}(k,k) & (\mathbf{a}\in\mathcal{A}_{10}) \end{cases}$$

transforms $\boldsymbol{\eta}(k,k)$ to an $(r_0 + r_1)$-dimensional probability vector $\boldsymbol{\eta}^{(0)}(k,k)$ without redundant elements. Note that, given the constraints, there is a one-to-one correspondence between $\boldsymbol{\eta}(k,k)$ and $\boldsymbol{\eta}^{(0)}(k,k)$. For estimating $\boldsymbol{\eta}(k,k)$ with a known color vector vector $\mathbf{x}$, the vector $(e_{\mathbf{a}}(k,k))_{\mathbf{a}\in\mathcal{A}'}$ defined in (8) is a sufficient statistic, whereas $\boldsymbol{\eta}^{(0)}(k,k)$ is the corresponding multinomial probability vector. Thus the uniform prior distribution for $\boldsymbol{\eta}$ consists of independent Dirichlet distributions $\mathrm{D}(1, \ldots, 1)$ in r dimensions for $\boldsymbol{\eta}(k,h)$ with $k < h$, and in $r_0 + r_1$ dimensions for $\boldsymbol{\eta}^{(0)}(k,k)$ for all $k \in \mathcal{C}$.

For Dirichlet prior distributions, well-known results on the Bayesian analysis of multinomially distributed data (see, e.g., Gelman et al. 1995) can be used to derive the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\eta})$ used in step 1 of the Gibbs sampler. If the prior distribution of $\boldsymbol{\theta}$ is Dirichlet with parameters T_k and the prior distribution of $\boldsymbol{\eta}(k,h)$ is Dirichlet with parameters $E_{\mathbf{a}}(k,h)$, while $\boldsymbol{\theta}$ and the $\boldsymbol{\eta}(k,h)$ are a priori independent, then the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\eta})$, given the complete

data $(\mathbf{y}, \mathbf{x})$, is given by independent Dirichlet distributions with parameters

$$(m_k + T_k)_{k \in \mathcal{C}} \quad \text{for } \boldsymbol{\theta},$$
$$(e_{\mathbf{a}}(k,h) + E_{\mathbf{a}}(k,h))_{\mathbf{a} \in \mathcal{A}} \quad \text{for } \boldsymbol{\eta}(k,h), 1 \le k < h \le c,$$

and

$$(e_{\mathbf{a}}(k,k) + E_{\mathbf{a}}(k,k))_{\mathbf{a} \in \mathcal{A}'} \quad \text{for } \boldsymbol{\eta}^{(0)}(k,k), 1 \le k \le c. \quad (22)$$

Step 1 of the Gibbs sampler is determined by these posterior distributions. As discussed earlier, for an invariant prior we use $T_k = 100c$ and $E_{\mathbf{a}}(k,h) = 1$ (all $k, h, \mathbf{a}$).

The relative frequency with which $X_i^{(p)} = X_j^{(p)}$ over a large number of runs after convergence of the Gibbs sampler is a good approximation of (18). Similarly, the average of $\eta_{\mathbf{a}}^{(p)}(X_i^{(p)}, X_j^{(p)})$ is a good approximation of (19). The posterior standard deviations can be used to indicate their precision.

If the matrix (18) indicates a clear block structure, then the vertices can be assigned to the classes (using an arbitrary color labeling), and (19), averaged over all pairs (i, j) with $X_i = h$, and $X_j = k$, can be used to make inferences about the probability distributions $\boldsymbol{\eta}$ of the relations between latent classes h and k. If there is no clear block structure, then it may be concluded that for the present number of colors, the model does not fit well.

5.1 Parameters for the Class Structure

We propose two parameters to give some insight into the adequacy of the obtained class structure. Their posterior means can be estimated directly from the Gibbs sampling results. The first is defined as

$$I_y = -\frac{2}{\sharp(\mathcal{N})} \sum_{(i,j) \in \mathcal{N}, i<j} \log(\eta_{y_{ij}}(X_i, X_j)). \quad (23)$$

This is, up to the multiplicative constant, the information in the observed relations, and also minus the log-likelihood, conditional on the class structure and on the probability parameter $\boldsymbol{\eta}$. This value is small if $\boldsymbol{\eta}$ and $\mathbf{X}$ determine to a large extent the observed relation pattern $\mathbf{y}$. For $c = 1$, the minimum value as a function of $\boldsymbol{\eta}$ is

$$I_y^{(1)} = -\sum_{\mathbf{a} \in \mathcal{A}'} p_y(\mathbf{a}) \log(p_y(\mathbf{a})), \quad (24)$$

where $p_y(\mathbf{a})$ denotes the fraction of pairs $(i, j) \in \mathcal{N}$ with $i < j$ and with $y_{ij} \in \{\mathbf{a}, \pi(\mathbf{a})\}$. The posterior mean of I_y may be greater or smaller than $I_y^{(1)}$; the fact that the posterior mean rather than the minimum is taken over $\boldsymbol{\eta}$ will make the value greater, but using more than one color will tend to make the value smaller. A class structure with $c \ge 2$ classes will be more satisfactory when the posterior mean of I_y is smaller.

The second parameter is

$$H_x = \frac{4}{n(n-1)} \sum_{i,j=1}^{n} \pi_{ij}(1 - \pi_{ij}), \quad (25)$$

where $\pi_{ij} = \Pr(X_i = X_j \mid \mathbf{y})$. This measures the extent to which the distribution of $\mathbf{X}$ defines one clear-cut partition of the vertices into classes. $H_x = 0$ if and only if the distribution of $\mathbf{X}$ is concentrated on one partition. If H_x is small, then it is relatively clear for every pair of vertices whether the vertices are in the same class [say, $P(X_i = X_j \mid \mathbf{y}) > .9$] or in different classes [say, $P(X_i = X_j \mid \mathbf{y}) < .1$]. The maximum value is 1, assumed when $\Pr(X_i = X_j \mid \mathbf{y}) = 1/2$ for all i, j.

The values of I_y and H_x can be helpful for assessing the adequacy of a model with a given number of colors, and for checking convergence of the Gibbs sampler, as discussed next.

5.2 Convergence Detection

The iteration steps of the Gibbs sampler are simple. Detecting convergence is not straightforward, however. This point is discussed, among others, by Gilks and Roberts (1996) and Gelman (1996). The reason is that the process converges not to a single value but to a stationary probability *distribution*. We propose for our model a measure for improving convergence and a measure for checking convergence.

The measure for improving convergence is to start the simulations with an overdispersed distribution for $(\boldsymbol{\theta}, \boldsymbol{\eta})$. This is meant to decrease the risk that the parameter vector is trapped in a suboptimal region. In our applications, we started the iterations with a large number M_0 (e.g., $M_0 = 5{,}000$) iterations, in which the prior distribution of $\boldsymbol{\theta}$ in the ith iteration is Dirichlet with all parameters equal to T_i^0, where T_i^0 decreases linearly from $T_1^0 = 10n$ to $T_{M_0}^0 = T$. (Recall that we used $T = 100c$ for the "real" prior.) Simultaneously, we multiplied the Dirichlet parameters $e_{\mathbf{a}}(k,h) + E_{\mathbf{a}}(k,h)$ in (22) (for $k < h$ as well as $k = h$) in the ith iteration by w_i, where w_i increases linearly from $w_1 = 1/n$ to $w_{M_0} = 1$, provided that the resulting number was not less than 1. These M_0 iterations were followed by another M_0 iterations according to the Dirichlet parameters in (22). Convergence is not assumed to have taken place before these $2M_0$ initial iterations.

To check convergence, we made multiple runs with independent starting points. Between each of these runs, after the $2M_0$ initial iterations, we compared moving averages of the "time series" of the values of (25) and (23) as a check of convergence. Whenever the values of these statistics and the estimated posterior means of relevant parameters [such as (18) and (19)] do not differ appreciably among the multiple sequences, we may assume that convergence has occurred.

A diagnostic for good mixing in single sequences, applicable if the prior distribution is invariant under relabeling of the colors, can be obtained by monitoring the vector $\mathbf{X}^{(p)}$ during the Gibbs sampling process. Because of the invariance of the posterior predictive distribution of $\mathbf{X}$, this distribution must be uniform. If, say, $c = 2$ and the data are such that the partition in two latent classes is reasonably clear but not too strong, then one will observe epochs during which some set of vertices tends to have color 1 and another set tends to have color 2, changing into epochs where colors 1 and 2 have changed places. Such color changes indicate good mixing as defined by Gilks and Roberts (1996), and thus can be taken as signs of convergence of the Gibbs sampler.

When one wishes to make comparisons between parallel simulation runs (cf. Gelman 1996), the parallel runs can also be compared on the basis of the colorings only. Drawing the new values of $(\boldsymbol{\theta}, \boldsymbol{\eta})$ in the Gibbs sampler depends only on the coloring and not on the previous value of $(\boldsymbol{\theta}, \boldsymbol{\eta})$, so that

the stochastic process $\mathbf{X}^{(p)}$ for $p=1,\ldots$ is a Markov chain, and it makes sense to monitor this process without considering the $(\boldsymbol{\theta}^{(p)}, \boldsymbol{\eta}^{(p)})$ process. To make comparisons between the within-sequence and between-sequence variability of the colorings, the partitions induced by colorings $\mathbf{X}^a$ and $\mathbf{X}^b$ can be compared using the symmetric difference distance proposed by Rand (1971) and studied by, for example, Lerman (1988) and Sibuya (1993). This distance is defined as the number of pairs (i,j) that are in the same class in one partition but in different classes in the other partition,

$$d(\mathbf{X}^a, \mathbf{X}^b) = \sum_{i,j=1}^{n} |I\{X_i^a = X_j^a\} - I\{X_i^b = X_j^b\}|.$$

Convergence may be assumed when the average between-sequence distances are not larger than the average within-sequence distances (cf. Gelman 1996).

In our experience, with n ranging between 10 and 80, convergence seems to be satisfactory after, say, $2M_0 = 10{,}000$ initial iterations, and often much earlier.

6. NONUNIFORM PRIOR DISTRIBUTIONS

If prior information is available, it can be incorporated into the prior distribution. A nonuniform prior distribution can identify the class labels and thereby circumvent the problems caused by nonidentifiability. The Gibbs sampler can then focus on the posterior means of the color-dependent probabilities $\boldsymbol{\eta}$ and the vertex color indicators $I\{X_i = k\}$, rather than on those of their invariant counterparts (17) and (15). We discuss various uses of nonuniform priors in this section.

6.1 Prior Class Probabilities

Prior distributions for the vertex colors can be used to identify the color labels. Suppose that prior knowledge is available for the class membership of a subclass $\mathcal{V}_p$ of the vertices. The basic model can then be changed by postulating for these vertices prior probability distributions

$$\Pr(X_i = k) = t_{ik}, i \in \mathcal{V}_p, k = 1, \ldots, c. \tag{26}$$

The other vertices still have probabilities (3), and all vertex colors are assumed to be independent. The Gibbs sampling procedure can be carried out as indicated earlier, with the only change that for the vertices in $\mathcal{V}_p$, the prior values θ_k in (21) are replaced by t_{ik} and the definition in (4) of the counts m_k used in (22) is replaced by

$$m_k = \sum_{\substack{i=1 \\ i \notin \mathcal{V}_p}}^{n} I(x_i = k).$$

An extreme instance is the case where it is assumed that a set of $c-1$ or c given vertices are all in different classes; for example, $\mathcal{V}_p = \{1, \ldots, c\}$, in which vertex i has color i, so that $t_{ii} = 1$ and $t_{ik} = 0$ for all $1 \le i \le c$ and all $k \ne i$. In a less extreme instance, one could postulate values

$$t_{ik} = \begin{cases} 1-\epsilon, & i = k \\ \epsilon/(c-1), & i \ne k \end{cases} \tag{27}$$

with a small positive ϵ (e.g., $\epsilon = .05$). In this way the possibility is left open for the data to override the prior ideas.

6.2 Confirmatory Blockmodeling

There may exist prior ideas about how the relations Y_{ij} should depend on the colors of vertices i and j. These can be expressed in prior distributions for the parameters $\boldsymbol{\eta}$. For some or all of the pairs (h,h) and (h,k), the vectors of probabilities $\boldsymbol{\eta}^{(0)}(h,h) = (\eta_{\mathbf{a}}^{(0)}(h,h))_{\mathbf{a} \in \mathcal{A}'}$ and $\boldsymbol{\eta}(h,k) = (\eta_{\mathbf{a}}(h,k))_{\mathbf{a} \in \mathcal{A}}$ $(h<k)$ can be given Dirichlet (or other) prior distributions that are not uniform. With respect to the determination of the parameters of these Dirichlet priors, it should be kept in mind that the sum of the parameters of a Dirichlet prior indicates the size of a sample that would influence the posterior as strongly as the prior.

As an example, Bock (1996a, Sec. 4) considers a model for cluster analysis, which is identical to our model in the case of a graph, $\mathcal{A} = \{(0,0),(1,1)\}$, where within-class relations are more likely to have the value (1,1) and between-class relations are more likely to have the value (0,0). Denoting $\eta(k,h) = \eta_{(1,1)}(k,h)$, this configuration can be achieved by using prior distributions of $\eta(k,k)$ concentrated on high values and priors of $\eta(k,h)$ for $h \ne k$ concentrated on low values. Specifically, $\eta(k,k)$ could be given the beta$(K, K/4)$ and $\eta(k,h)$ the beta$(K/4, K)$ distribution, where K is less than the number of observed relations that are informative for estimating $\eta(k,h)$; for example, $K = n^2/10c^2$ (if class sizes are not too strongly different).

In the analysis of social networks, the analyst may have a hypothesis about the pattern of ties between actors depending on their positions. This can be expressed as prior ideas about the probability distributions $\boldsymbol{\eta}(h,k)$ of the relations between some of the positions h and k. White et al., (1976) first described such relational structures in terms of blockmodel image matrices for a two-position model. (The image matrix is the deterministic analog of our parameter $\boldsymbol{\eta}$.) More complex systems with three or more positions also have been studied; some such structures were presented by Wasserman and Faust (1994, sec. 10.3.3). In their terminology, the structure mentioned in the preceding paragraph is a cohesive subgroups structure. In a center-periphery structure, one class has many relations internally and with all others, whereas all other classes have few relations both internally and with each other. In a hierarchical structure for directed graphs, the classes are ordered and there are especially many relations from a given class directed to the immediately higher class (but not reversely). (More examples were given in Wasserman and Faust 1994.) It can be investigated whether such a block structure exists in an observed social network by carrying out the Gibbs sampling analysis using nonuniform prior Dirichlet distributions for the probabilities $\boldsymbol{\eta}(h,k)$ of the relations between some pairs of groups (h,k), whereas the priors are uniform for the pairs of groups for which there is no theoretical expectation.

7. EXAMPLE: KAPFERER'S TAILOR SHOP

Kapferer (1972) observed interactions in a tailor shop in Zambia over a period of 10 months, focusing on the changing patterns of alliance among $n=39$ workers during extended negotiations for higher wages. The data relate to two types of interaction: work- and assistance-related relationships

(a nondirected relation) and friendship interactions (a directed relation). Data were recorded on two occasions between which an abortive strike occurred; these are available in *Ucinet* 5 (Borgatti, Everett, and Freeman 1998).

The approach applied by Kapferer was based on dividing the workers into 3 classes by rearranging the columns and rows of the adjacency matrix, a procedure devised by Beum and Brundage (1950). This ad hoc method was aimed at defining classes consisting of individuals whose interactions take place mainly within the class, and may be regarded as a rough form of blockmodeling, directed at finding cohesive classes.

We use the rank orders of the vertices in the *Ucinet* dataset as the vertex labels. The employees' occupational categories were, in the order of their prestige according to a survey among employees (Kapferer 1972, p. 55), head tailor (worker nr 19), cutter {16}, line 1 tailor {1–3, 5–7, 9, 11–14, 21, 24}, button machiner {25–26}, line 3 tailor {8, 15, 20, 22–23, 27–28}, ironer {29, 33, 39}, and cotton boy {30–32, 34–38}. An additional category was line 2 tailor, which included employees {4, 10, 17–18}.

To illustrate our model, we analyze the poststrike relations. The network densities were .30 for assistance and .20 for friendship. These two ties were positively associated, with log odds ratio 1.11. The assistance and friendship relations were analyzed simultaneously. This yielded a relational structure with $R=4$ possible relations from one vertex to another:

$$a_1 = 0, \text{ no tie}$$
$$a_2 = A, \text{ assistance, no friendship}$$
$$a_3 = F, \text{ friendship, no assistance}$$
$$a_4 = AF, \text{ assistance and friendship.}$$

Because assistance is nondirected and friendship is directed, the alphabet $\mathcal{A}$ consists of the $r_0=4$ symmetric relations

$$\mathcal{A}_0 = \{(0,0),(A,A),(F,F),(AF,AF)\}$$

and the $2r_1=4$ asymmetric relations

$$\mathcal{A}_1 = \{(0,F),(A,AF),(F,0),(AF,A)\}.$$

Thus the alphabet has a total of $r=8$ dyadic relations. Their frequencies are presented in Table 1. The upper diagonal part of this table is defined as $\mathcal{A}'$; the total frequency in this part of the table is $\binom{39}{2}=741$. This table shows, for example, that the majority of the dyads, (698 out of 741) have an asymmetric relation, and reciprocated friendship tends to go along with assistance (viz., in 46 out of 52 cases).

We ran the Gibbs sampler in the aforementioned (unidentified) specification, the prior Dirichlet distribution for the vertex color probabilities with parameter value $T=100c$. After $2M_0$ = 50,000 iterations, we assumed convergence; subsequently, we carried out 50,000 iterations to estimate the posterior distributions. For each of the values $2 \le c \le 5$, we executed three independent Gibbs samplers. These three gave very similar results, and there was no reason to doubt that convergence was satisfactory long before the 50,000 warming-up iterations were concluded.

Table 1. Frequencies of Dyadic Ties $(a_t, a_v) \in \mathcal{A}$

	a_v			
a_t	0	*A*	*F*	*AF*
0	493		19	
A		153		24
F	19		6	
AF		24		46

The values of the information I_y and the parameter H_x, expressing the clearness of the block structure, are given in Table 2. (In all cases, the differences between the I_y values and between the H_x values produced in the three parallel runs were less than .01.) This table shows that the information I_y contained in the latent class structure about the probability distribution of the relations decreases when c goes from 2 to 4, and changes little when c increases further to 5. In addition, the clearest class structure, as expressed by H_x, is obtained for $c=3$.

For $c=3$, Table 3 represents the matrix of posterior means (18) of the probability that two vertices have the same color. (The three replications of the Gibbs sampler gave practically the same values for all parameter estimates.) The numbers given represent the rounded first digit after the decimal point, but with the number 9 referring to values from .85 up to 1.0. The vertices were reordered to obtain a clear block structure. It is clear that the vertices 1–3, 5, 7, 12–14, 16, 19, 24, 25, and (less evidently) 30 and 34 tend to belong to the same class, (now labeled class 1), and vertices 4, 6, 8, 10, 15, 17, 18, 20, 22, 23, 26–29, 31–33, and 35–39 also belong to one class (class 2). Vertex 11 tends to be a class of its own (class 3). Vertices 9 and 21 occupy intermediate positions. The description of the functions of the 39 actors shows that class 1 consists mainly of higher-prestige workers, with the two cotton boys, (30 and 34) as the exceptions (but their belonging to class 1 is a bit doubtful). Kapferer described how vertex 11, a line 1 tailor named Lyashi, gradually emerged as a major challenger to the supervisors for power and leadership. Although not fully successful, he achieved an important and unique position in the group, which is clearly revealed in our analysis.

A good representation of the relations between the three classes is obtained by considering the average posterior probabilities (19) between the vertices comprised in these three classes, disregarding vertices 9 and 21. The result is displayed in Table 4.

The main distinctions between the classes are as follows. Within class 1, the posterior probability is .38 for assistance without friendship and .16 for assistance with mutual friend-

Table 2. Parameters for the Class Structure for Kapferer's Dataset

c	I_y	H_x
2	.94	.24
3	.91	.21
4	.89	.26
5	.89	.27

Table 3. Estimated Posterior Probabilities That Two Vertices Have the Same Color

i	Posterior probability $P(X_i = X_j \mid \mathbf{y})$																																						
1	—	9	9	9	9	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	9	—	9	9	9	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3	9	9	—	9	9	6	9	9	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	9	9	9	—	9	6	9	9	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7	9	9	9	9	—	6	9	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	6	6	6	6	6	—	6	6	6	5	5	7	6	4	5	5	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0
12	9	9	9	9	9	6	—	9	9	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
13	9	9	9	9	9	6	9	—	9	9	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14	9	9	9	9	9	6	9	9	—	8	9	6	9	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
16	8	8	9	9	8	5	8	9	8	—	9	5	9	9	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
19	9	9	9	9	9	5	9	9	9	9	—	5	9	9	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
21	6	6	6	6	6	7	6	6	6	5	5	—	6	4	7	6	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	0
24	9	9	9	9	9	6	9	9	9	9	9	6	—	8	8	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
25	8	8	8	8	8	4	8	8	8	9	9	4	8	—	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
30	8	8	8	8	8	5	8	8	8	7	7	7	8	7	—	8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0
34	7	7	7	7	7	5	7	7	7	6	6	6	7	6	8	—	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0
4	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
6	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
8	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
10	1	1	1	1	1	5	1	1	1	0	0	4	1	0	2	3	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
15	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
17	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
18	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
20	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0
22	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	9	0
23	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	9	0
26	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	9	0
27	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	9	0
28	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	9	0
29	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	9	0
31	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	9	0
32	1	1	1	1	1	4	1	1	1	1	1	4	1	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	9	0
33	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	9	0
35	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	9	0
36	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	9	0
37	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	9	0
38	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	9	0
39	0	0	0	0	0	4	0	0	0	0	0	4	0	0	2	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	—	0
11	0	0	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—

NOTE: Digits given are the first digit after the decimal point. Rows and columns are similarly ordered.

Table 4. Estimated Posterior Probabilities $E(\eta_a(X_i, X_j) \mid \mathbf{y})$ of Relations, Averaged Over Vertices *i* in Class *h* and Vertices *j* in Class *k*

a	*h*	*k* = 1	*k* = 2	*k* = 3
(0,0)	1	.27	.72	.14
	2	.72	.78	.31
	3	.14	.31	
(*A*,*A*)	1	.38	.19	.15
	2	.19	.15	.22
	3	.15	.22	
(*F*,*F*)	1	.03	.01	.07
	2	.01	.01	.04
	3	.07	.04	
(*AF*,*AF*)	1	.16	.03	.42
	2	.03	.05	.07
	3	.42	.07	
(0,*F*)	1	.01	.00	.05
	2	.03	.00	.27
	3	.04	.03	
(*A*,*AF*)	1	.07	.01	.08
	2	.02	.00	.04
	3	.05	.03	

ship; these probabilities are much lower within class 2 and also between classes 1 and 2. Class 3 (the single vertex 11) has a .42 posterior probability of assistance with mutual friendship with class 1. Thus workers in class 1, mainly higher-prestige workers, have more assistance and friendship relations among each other. Actor 11 presents a unique pattern of friendship ties and occupies a particularly central position, thus playing a crucial role in connecting the network. Compared to classes 1 and 2, he has the highest estimated probability of assistance and of reciprocated friendship ties with class 1 and also (albeit at a lower level) with class 2. In addition, he extends a relatively large number of unreciprocated friendship ties to actors in class 2. He seems to occupy a special position in his propensity to express friendship ties.

When the Gibbs sampler is run with $c = 4$ colors, the difference in the resulting class structure is that a new class, consisting of vertices 16, 19, and 25, splits off from class 1; these vertices tend to have slightly more ties of the types (A, A) with each other and with the other vertices of class 1. For $c = 4$, there are more vertices with an ambiguous position, as reflected by the higher value for H_x. In particular, the lat-

ter property makes the results for 4 colors less attractive to interpret.

8. DISCUSSION

This article has presented a statistical approach to posterior blockmodeling of (valued) graphs and digraphs. In our experience, the algorithm works well for graphs of up to at least 100 vertices. A program for carrying out these analyses is available at http://stat.gamma.rug.nl/stocnet/.

Our analysis is based on a model that assumes that the probability distribution of the relation between two vertices depends only on the latent classes to which the vertices belong and that the relations are independent conditionally on these classes. This is analogous to usual latent class models. Although it may be considered rather simplistic as a probability model for graphs as observed, for example, in the social sciences, this model is a step forward with respect to the deterministic models for blockmodeling and the rather restrictive p_1 blockmodel proposed by Wasserman and Anderson (1987). The restriction of inference to invariant functions of the parameters gives a practical and elegant solution to the problems posed by the nonidentifiability of the class labels.

Several extensions of the model might be worthwhile to develop. One possibility is to extend the model with observed covariates (defined at the level of the vertex or, more generally, of the ordered pair (i, j)). A second possibility is to adopt more complex conditional dependence assumptions for entries in $\mathbf{Y}$ conditional on the latent classes $\mathbf{X}$. The structural parameters contained in the p^* model proposed by Wasserman and Pattison (1996) provide a natural point of departure for such an analysis. A model combining latent classes of vertices with the structural parameters of the p^* model seems feasible in principle, but the elaboration of statistical procedures for such a model will require considerable effort.

Another possibility could be to drop the assumption that the colors X_i are iid, and replace this by a more complex model for $\mathbf{X}$, expressing "emergent" social roles such as leader/follower or center/periphery. This seems useful especially if the data are not restricted (as they are in this article) to a single observation on a relational structure, but rather include repeated measures or a sample of such observations.

From the standpoint of practical data analysis of relational data, our a posteriori blockmodeling has a number of advantages over current blockmodeling methods as reviewed by, for example, Wasserman and Faust (1994). It is based on an explicit probability model and does not make any restrictions about the probability distributions of the relations between the latent classes. Further, it does not force all vertices into the c hypothesized classes but, through the matrix of posterior probabilities (18) that two vertices belong to the same class, expresses the class structure in a way that allows some of the vertices to have an equivocal position. This is often realistic when fitting this model to a given dataset. Finally, the facts that the method can be used not only for dichotomous relations, but also for any finite set of possible relations, and that it allows missing and structurally predetermined relations, adds to its quite general applicability.

[Received January 2000. Revised January 2001.]

REFERENCES

Anderson, C. J., Wasserman, S., and Faust, K. (1992), "Building Stochastic Blockmodels," *Social Networks*, 14, 137–161.

Beum, C. O., and Brundage, E. G. (1950), "A Method for Analysing the Sociomatrix," *Sociometry*, 12, 141–145.

Bock, H. H. (1996a), "Probabilistic Models in Cluster Analysis," *Computational Statistics and Data Analysis*, 23, 5–28.

——— (1996b), "Probability Models and Hypotheses Testing in Partitioning Cluster Analysis," in *Clustering and Classification*, eds. P. Arabie, L. J. Hubert, and G. de Soete, River Edge, NJ: World Scientific Publishing, pp. 377–453.

Borgatti, S., Everett, M. G., and Freeman, L. C. (1998), *UCINET V Reference Manual*, Columbia, SC: Analytic Technologies.

Faust, K. (1988), "Comparison of Methods for Positional Analysis: Structural and General Equivalences," *Social Networks*, 10, 313–341.

Fienberg, S. E., and Wasserman, S. (1981), "Categorical Data Analysis of Single Sociometric Relations," in *Sociological Methodology–1981*, ed. S. Leinhardt San Francisco: Jossey-Bass, pp. 156–192.

Frank, O. (1988a), "Triad Count Statistics," *Discrete Mathematics*, 72, 141–149.

——— (1988b), "Random Sampling and Social Networks: A Survey of Various Approaches," *Mathématiques, Informatique et Sciences Humaines*, 26(104), 19–33.

Frank, O., and Harary, F. (1982), "Cluster Inference by Using Transitivity Indices in Empirical Graphs," *Journal of the American Statistical Association*, 81, 835–840.

Gelman, A. (1996), "Inference and Monitoring Convergence," in *Markov Chain Monte Carlo in Practice*, eds. W. Gilks, S. Richardson, and D. Spiegelhalter, London: Chapman and Hall, pp. 131–144.

Gelman, A., Carlin, J. B., Stern, H., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman and Hall.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences" (with discussion), *Statistical Science*, 7, 457–511.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Gilks, W. R., and Roberts, G. O. (1996), "Strategies for Improving MCMC," in *Markov Chain Monte Carlo in Practice*, eds. W. Gilks, S. Richardson, and D. Spiegelhalter, London: Chapman and Hall, pp. 89–114.

Holland, P., Laskey, K. B., and Leinhardt, S. (1983), "Stochastic Blockmodels: Some First Steps," *Social Networks*, 5, 109–137.

Holland, P., and Leinhardt, S. (1981), "An Exponential Family of Probability Distributions for Directed Graphs," *Journal of the American Statistical Association*, 76, 33–50.

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.

Kapferer, B. (1972), *Strategy and Transaction in an African Factory*, Manchester, U.K.: Manchester University Press.

Lerman, I. C. (1988), "Comparing Partitions (Mathematical and Statistical Aspects)," in *Classification and Related Methods of Data Analysis*, ed. H. H. Block, Amsterdam: Elsevier/North-Holland, pp. 121–131.

Lorrain, F., and White, H. C. (1971), "Structural Equivalence of Individuals in Social Networks," *Journal of Mathematical Sociology*, 1, 49–80.

Rand, W. M. (1971), "Objective Criteria for the Comparison of Clustering Methods," *Journal of the American Statistical Association*, 66, 846–850.

Robert, C. P. (1996), "Mixtures of Distributions: Inference and Estimation," in *Markov Chain Monte Carlo in Practice*, eds. W. Gilks, S. Richardson, and D. Spiegelhalter, London: Chapman and Hall, pp. 441–464.

Sibuya, M. (1993), "Random Partition of a Finite Set by Cycles of Permutation," *Japan Journal of Industrial and Applied Mathematics*, 10, 69–84.

Snijders, T., and Nowicki, K. (1997), "Estimation and Prediction for Stochastic Blockmodels for Graphs With Latent Block Structure," *Journal of Classification*, 14, 75–100.

Wasserman, S., and Anderson, C. (1987), "Stochastic a Posteriori Blockmodels: Construction and Assessment," *Social Networks*, 9, 1–36.

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

Wasserman, S., and Pattison, P. (1996), "Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^*," *Psychometrika*, 61, 401–425.

Wellman, B., Frank, O., Espinoza, V., Lundquist, S., and Wilson, C. (1991), "Integrating Individual, Relational and Structural Analysis," *Social Networks*, 13, 223–249.

White, H. C., Boorman, S., and Breiger, R. L. (1976), "Social Structure From Multiple Networks," *American Journal of Sociology*, 81, 730–780.