J. R. Statist. Soc. A (2007)
170, *Part* 2, *pp.* 1–22

Model-based clustering for social networks

Mark S. Handcock and Adrian E. Raftery

University of Washington, Seattle, USA

and Jeremy M. Tantrum

Microsoft adCenter Laboratories, USA

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 11th, 2006*, Professor T. J. Sweeting *in the Chair*]

Summary. Network models are widely used to represent relations between interacting units or actors. Network data often exhibit transitivity, meaning that two actors that have ties to a third actor are more likely to be tied than actors that do not, homophily by attributes of the actors or dyads and clustering. Interest often focuses on finding clusters of actors or ties, and the number of groups in the data is typically unknown. We propose a new model, the *latent position cluster model*, under which the probability of a tie between two actors depends on the distance between them in an unobserved Euclidean 'social space', and the actors' locations in the latent social space arise from a mixture of distributions, each corresponding to a cluster. We propose two estimation methods: a two-stage maximum likelihood method and a fully Bayesian method that uses Markov chain Monte Carlo sampling. The former is quicker and simpler, but the latter performs better. We also propose a Bayesian way of determining the number of clusters that are present by using approximate conditional Bayes factors. Our model represents transitivity, homophily by attributes and clustering simultaneously and does not require the number of clusters to be known. The model makes it easy to simulate realistic networks with clustering, which are potentially useful as inputs to models of more complex systems of which the network is part, such as epidemic models of infectious disease. We apply the model to two networks of social relations. A free software package in the R statistical language, latentnet, is available to analyse data by using the model.

Keywords: Bayes factor; Dyad; Latent space; Markov chain Monte Carlo methods; Mixture model; Transitivity

1. Introduction

Networks are widely used to represent data on relations between interacting actors or nodes. They can be used to describe the behaviour of epidemics, the interconnectedness of corporate boards, networks of genetic regulatory interactions and computer networks, among others. In social networks, each actor represents a person or social group, and each link, tie or arc represents the presence or strength of a relationship between two actors. Nodes can be used to represent larger social units (groups, families or organizations), objects (airports, servers or locations) or abstract entities (concepts, texts, tasks or random variables).

Social network data typically consist of a set of n actors and a relational tie $y_{i,j}$, measured on each ordered pair of actors $i, j = 1, \ldots, n$. In the simplest cases, $y_{i,j}$ is a dichotomous variable, indicating the presence or absence of a relation of interest, such as friendship, collaboration or transmission of information or disease. The data are often represented by an $n \times n$ sociomatrix

Address for correspondence: Adrian E. Raftery, Center for Statistics and the Social Sciences, University of Washington, Box 354320, Seattle, WA 98195-4320, USA.
E-mail: raftery@u.washington.edu

 0964–1998/07/170000

Y. In the case of binary relations, the data can also be thought of as a graph in which the nodes are actors and the (directed) edges are $\{(i, j) : y_{i,j} = 1\}$. When (i, j) is an edge we write $i \to j$.

A feature of most social networks is transitivity of relations whereby two actors that have ties to a third actor are more likely to be tied than actors that do not. Transitivity has been extensively studied both empirically and theoretically (White *et al.*, 1976). Transitivity can lead to some clustering of relationships within the network.

The likelihood of a tie usually depends on attributes of the actors. For example, for most social relations the likelihood of a relationship is a function of the age, gender, geography, race and status of the individuals. In addition, ties are often more likely to occur between actors that have similar attributes than between those who do not, a tendency that we call homophily by attributes (Lazarsfeld and Merton, 1954; Freeman, 1996; McPherson *et al.*, 2001). Although homophily by attributes usually implies increased probability of a tie, the effect may be reversed (e.g. gender and sexual relationships).

Many social networks exhibit clustering beyond what can be explained by transitivity and homophily on observed attributes. This can be driven by homophily on unobserved attributes or on endogenous attributes such as position in the network (Wasserman and Faust, 1994), 'self-organization' into groups or a preference for popular actors. Often the key questions in a social network analysis revolve around the identification of clusters, but conclusions about clustering are usually drawn by informal visual examinations of the network rather than by more formal inference methods (Liotta, 2004).

Existing stochastic models struggle to represent the three common features of social networks that we have mentioned, namely transitivity, homophily by attributes and clustering. Holland and Leinhardt (1981) proposed a model in which each dyad—by which we mean each pair of actors—had ties independently of every other dyad. This model was inadequate because it did not capture any of the three characteristics. Frank and Strauss (1986) generalized it to the case in which dyads exhibit a form of Markovian dependence: two dyads are dependent, conditional on the rest of the graph, only when they share an actor. This can represent transitivity, although not the other two characteristics. Exponential random graph models generalize this idea further and can represent some forms of transitivity (Snijders *et al.*, 2005).

Models based only on the distribution of the number of edges linking to the actors, or degree distribution, are popular in physics and applied mathematics; for a review see Newman (2003). These are also quite restrictive and often do not model any of the three key features of network data that we have mentioned (Snijders, 1991).

The seminal work on structural equivalence by Lorrain and White (1971) motivated statistical procedures for clustering or 'blocking' relational data (*block models*). Blocking consists of a known prespecified partition of the actors into discrete blocks and, for each pair of blocks, a statement of the presence or absence of a tie within or between the blocks. This requires knowledge of the partition, which will often not be available. Breiger *et al.* (1975) and White *et al.* (1976) developed and compared alternative algorithms. Subsequent work in this area has been on deterministic algorithms to block actors into prespecified theoretical types (Doreian *et al.*, 2005). Here we focus on stochastic models for networks, which seem more appropriate for many applications.

Fienberg and Wasserman (1981) developed a probabilistic model for structural equivalence of actors in a network, under which the probabilities of relationships with all other actors are the same for all actors in the same class. This can be viewed as a stochastic version of a block model. It can represent clustering, but only when the cluster memberships are known. Wasserman and Anderson (1987) and Snijders and Nowicki (1997) extended these models to latent classes; the difference is that these latent class models do not assume cluster memberships to

be known, but instead estimate them from the data. Nowicki and Snijders (2001) presented a model where the number of classes is arbitrary and unknown. The model assumes that the probability distribution of the relation between two actors depends only on the latent classes to which the two actors belong and the relations are independent conditionally on these classes. These models do capture some kinds of clustering, but they do not represent transitivity within clusters or homophily on attributes. Tallberg (2005) extended this model to represent homophily on observed attributes.

The idea of representing a social network by assigning positions in a continuous space to the actors was introduced in the 1970s; see, for example, McFarland and Brown (1973) and Breiger *et al.* (1975), who used multidimensional scaling to do this, and this approach has been widely used since (Wasserman and Faust, 1994). A strength of this approach is that it takes account of transitivity automatically and in a natural way. A disadvantage is that a dissimilarity measure must be supplied to the algorithm for each dyad, and many different dissimilarity measures are possible, so the results depend on a choice for which there is no clear theoretical guidance.

The latent space model of Hoff *et al.* (2002) is a stochastic model of the network in which each actor has a latent position in a Euclidean space, and the latent positions are estimated by using standard statistical principles; thus no arbitrary choice of dissimilarity is required. This model automatically represents transitivity and can also take account of homophily on observed attributes in a natural way. This approach was applied to international relations networks by Hoff and Ward (2004) and was extended to include random actor-specific effects by Hoff (2005). A similar model was proposed by Schweinberger and Snijders (2003), but using an ultrametric space rather than a Euclidean space.

Here we propose a new model, the latent position cluster model, that takes account of transitivity, homophily on attributes and clustering simultaneously in a natural way. It extends the latent space model of Hoff *et al.* (2002) to take account of clustering, using the ideas of model-based clustering (Fraley and Raftery, 2002). The resulting model can be viewed as a stochastic block model with transitivity within blocks and homophily on attributes. It can also be viewed as a generalization of latent class models to allow heterogeneity of structure within the classes.

In Section 2 we describe the latent position cluster model. In Section 3 we give two different ways of estimating it. One is a two-stage maximum likelihood estimation method, which is relatively fast and simple. The other is a fully Bayesian method that uses Markov chain Monte Carlo (MCMC) sampling; this is more complicated but performs better in our examples. In Section 4 we propose a Bayesian approach to choosing the number of groups in the data by using approximate conditional Bayes factors. In Section 5 we illustrate the method by using two social network data sets.

2. The latent position cluster model for social networks

The data that we model consist of an $n \times n$ sociomatrix Y, with entries $y_{i,j}$ denoting the value of the relation from actor i to actor j, possibly in addition to covariate information $X = \{X_{i,j}\}$. We focus on binary-valued relations, although the methods in this paper can be extended to more general relational data. Both directed and undirected relations can be analysed with our methods, although the models are slightly different in the two cases.

We assume that each actor has an unobserved position in a d-dimensional Euclidean latent social space, as in Hoff *et al.* (2002). We then assume that the presence or absence of a tie between two individuals is independent of all other ties, given the positions $Z = \{z_i\}$ in social space of the two individuals. Thus

$$P(Y|Z, X, \beta) = \prod_{i \neq j} P(y_{i,j}|z_i, z_j, x_{i,j}, \beta), \tag{1}$$

where $X = \{x_{i,j}\}$ denotes observed characteristics that may be dyad specific and vector valued, and β denotes parameters to be estimated. We model $P(y_{i,j}|z_i, z_j, x_{i,j}, \beta)$ by using a logistic regression model in which the probability of a tie depends on the Euclidean distance between z_i and z_j in social space:

$$\text{log-odds}(y_{i,j} = 1|z_i, z_j, x_{i,j}, \beta) = \beta_0^{\mathrm{T}} x_{i,j} - \beta_1 |z_i - z_j|, \tag{2}$$

where the log-odds of an event A is $\text{log-odds}(A) = \log[P(A)/\{1 - P(A)\}]$. The model accounts for transitivity, through the latent space, as well as homophily on the observed attributes X. To identify the scale of the positions and β_0 and β_1, we restrict the positions to have unit root mean square:

$$\sqrt{\left(\frac{1}{n}\sum_i |z_i|^2\right)} = 1.$$

To represent clustering, we assume that the z_is are drawn from a finite mixture of G multivariate normal distributions, each representing a different group of actors. Each multivariate normal distribution has a different mean vector, and a spherical covariance matrix, with variances that differ between groups. Thus

$$z_i \sim \sum_{g=1}^{G} \lambda_g \, \text{MVN}_d(\mu_g, \sigma_g^2 I_d), \tag{3}$$

where λ_g is the probability that an actor belongs to the gth group, so that $\lambda_g \geqslant 0 \, (g = 1, \ldots, G)$ and $\Sigma_{g=1}^{G} \lambda_g = 1$, and I_d is the $d \times d$ identity matrix. The choice of spherical covariance matrices is motivated by the fact that the likelihood is invariant to rotations of the latent social space, so it seems reasonable that the model be specified independently of the co-ordinate system. Model (3) was proposed as a model for clustering of observed variables by Banfield and Raftery (1993).

3. Estimation

We propose two different estimation methods for the latent position cluster model. The first is a two-stage method that first computes the maximum likelihood estimator of the (non-clustering) latent space model, and then computes the maximum likelihood estimator for the mixture model applied to the resulting estimated latent positions. This is fast and relatively simple, but it does not take advantage of the clustering information when estimating the latent positions. The second method is fully Bayesian and uses MCMC sampling; it estimates the latent positions and the clustering model simultaneously. This is more demanding computationally and algebraically than the first method, but it performs better in our examples.

3.1. Two-stage maximum likelihood estimation

The first stage is to carry out maximum likelihood estimation of the latent positions for the (non-clustering) latent space model of Hoff *et al.* (2002), as described there. This is fairly straightforward because the log-likelihood is convex as a function of the distances between actors, although not as a function of the actors' positions. One can thus rapidly find estimates of the distances, and then find a set of latent positions that approximate them by multidimensional scaling. This gives a good starting-point for a non-linear optimization method.

The second stage is to find a maximum likelihood estimator of the mixture model conditionally on the latent positions that are estimated at the first stage. This can be done by using the EM algorithm (Dempster *et al.*, 1977). It has been implemented for model (3) in a clustering context in the R package `mclust` (Fraley and Raftery, 1998, 2002, 2003). The likelihood function for model (3) does not have a unique local maximum, and the local maximum that is found by the EM algorithm can depend on the starting values. Here we use starting values from hierarchical model-based clustering (Banfield and Raftery, 1993).

This estimation method is fast and simple, and yields a close match between the estimated latent positions and cluster memberships. However, by not estimating the latent positions and the cluster model at the same time, we lose information from the cluster structure that may be useful in estimating the latent positions, and we lose information on the uncertainty about the latent positions that can be useful in clustering. We now describe a simultaneous estimation method that does not have these disadvantages.

3.2. Bayesian estimation

Our second method consists of fully Bayesian estimation of the latent position cluster model given by equations (1)–(3), using MCMC sampling. We introduce the new variables K_i, equal to g if the ith actor belongs to the gth group, as is standard in Bayesian estimation of mixture models (e.g. Diebolt and Robert (1994)).

We specify prior distributions for the parameters $\beta = (\beta_0^{\mathrm{T}}, \beta_1)^{\mathrm{T}}$, $\lambda = (\lambda_1, \ldots, \lambda_G)$, σ_g^2 and μ_g, as follows:

$$\beta \sim \mathrm{MVN}_p(\xi, \Psi),$$

$$\lambda \sim \mathrm{Dirichlet}(\nu),$$

$$\sigma_g^2 \overset{\mathrm{IID}}{\sim} \sigma_0^2 \, \mathrm{Inv}\chi_\alpha^2, \qquad g = 1, \ldots, G,$$

$$\mu_g \overset{\mathrm{IID}}{\sim} \mathrm{MVN}_d(0, \omega^2 I_d), \qquad g = 1, \ldots, G,$$

where ξ, Ψ, $\nu = (\nu_1, \ldots, \nu_G)$, σ_0^2, α and ω are hyperparameters to be specified by the user.

We set $\nu_g = 3$, which puts low probability on small group sizes, and $\xi = 0$ and $\Psi = 2I$, which allow a wide range of values of β. We take $\alpha = 2$ and $\sigma_0^2 = 0.103$ (the fifth percentile of the χ_2^2-distribution), which implies a prior density on σ_g^2 that has 90% of its mass between 0.017 and 1, corresponding to groups whose standard deviation can be as small as 13% of the average radius of the data. Finally, we specify $\omega^2 = 2$, which ensures that the prior density of the means is relatively flat over the range of the data.

Our MCMC algorithm iterates over the model parameters with the priors given above, the latent positions z_i and the group memberships K_i. Where possible we sample from the full conditional posterior distributions as in Gibbs sampling; otherwise we use Metropolis–Hastings steps. Let 'others' denote those of the parameters, latent positions and group memberships that are not explicitly specified in the following formulae. The full conditional posterior distributions are

$$z_i | K_i = g, \text{others} \propto \phi_d(z_i; \mu_g, \sigma_g^2 I_d) \; P(Y|Z, X, \beta), \qquad i = 1, \ldots, n, \tag{4}$$

$$\beta | Z, \text{others} \propto \phi_p(\beta; \xi, \Psi) \; P(Y|Z, X, \beta), \tag{5}$$

$$\lambda | \text{others} \sim \mathrm{Dirichlet}(m + \nu), \tag{6}$$

$$\mu_g|\text{others} \sim \text{MVN}_d\left(\frac{m_g \bar{z}_g}{m_g + \sigma_g^2/\omega^2}, \frac{\sigma_g^2}{m_g + \sigma_g^2/\omega^2} I\right), \qquad g = 1, \ldots, G, \tag{7}$$

$$\sigma_g^2|\text{others} \sim (\sigma_0^2 + d s_g^2)\, \text{Inv}\chi^2_{\alpha + m_g d}, \qquad g = 1, \ldots, G, \tag{8}$$

$$P(K_i = g|\text{others}) = \frac{\lambda_g\, \phi_d(z_i; \mu_g, \sigma_g^2 I_d)}{\sum_{r=1}^{G} \lambda_r\, \phi_d(z_i; \mu_r, \sigma_r^2 I_d)}, \qquad i = 1, \ldots, n,\ g = 1, \ldots, G, \tag{9}$$

where

$$m_g = \sum_{i=1}^{n} I_{[K_i = g]},$$

$$s_g^2 = \frac{1}{d} \sum_{i=1}^{n} (z_i - \mu_g)^{\mathrm{T}} (z_i - \mu_g) I_{[K_i = g]},$$

$$\bar{z}_g = \frac{1}{m_g} \sum_{i=1}^{n} z_i I_{[K_i = g]},$$

and $\phi_d(\cdot; \mu, \Sigma)$ is the d-dimensional multivariate normal density.

Our algorithm is then as follows.

Step 1: use Metropolis–Hastings steps to sample Z_{t+1}, updating each actor in random order.

(a) Propose $Z_i^* \sim \text{MVN}_d(Z_{ti}, \delta_Z^2 I_d)$.
(b) With probability equal to

$$\frac{P(Y|Z^*, X, \beta_t)\, \phi_d(Z_i^*; \mu_{K_i}, \sigma_{K_i}^2 I_d)}{P(Y|Z_t, X, \beta_t)\, \phi_d(Z_{it}; \mu_{K_i}, \sigma_{K_i}^2 I_d)},$$

set the ith element of Z_{t+1} to Z_i^*. Otherwise set it to Z_{it}.

Step 2: use Metropolis–Hastings steps to sample β_{t+1}.

(a) Propose $\beta^* \sim \text{MVN}_d(\beta_t, \delta_\beta^2 I_p)$.
(b) With probability equal to

$$\frac{P(Y|Z_{t+1}, X, \beta^*)\, \phi_p(\beta^*; \xi, \Psi)}{P(Y|Z_{t+1}, X, \beta_t)\, \phi_p(\beta_t; \xi, \Psi)},$$

set $\beta_{t+1} = \beta^*$. Otherwise set $\beta_{t+1} = \beta_t$.

Step 3: update K_i, μ_g, σ_g^2 and λ_g from expressions (6)–(9).

The proposal distribution variance parameters, δ_Z and δ_β, are set by the user to achieve good performance of the algorithm. On the basis of some experimentation, we used $\delta_Z = 10$ and $\delta_\beta = 0.5$.

3.3. Identifiability of positions and cluster labels

As the likelihood is a function of the latent positions only through their distances, it is invariant to reflections, rotations and translations of the latent positions. The likelihood is also invariant

to relabelling of the clusters, in the sense that permuting the cluster labels does not change the likelihood. This is often referred to as the label switching problem (Stephens, 2000).

We resolve these non-identifiabilities (or near non-identifiabilities in the Bayesian context) by post-processing the MCMC output. One simple two-stage approach to this would be as follows. First carry out a Procrustes transformation (Sibson, 1979) of each posterior draw of the latent positions to resolve the invariance to reflections, rotations and translations, following Oh and Raftery (2001) and Hoff *et al.* (2002). The target configuration would be the positions that are produced by the two-stage maximum likelihood procedure. Second, use the relabelling algorithm of Celeux *et al.* (2000) to solve the label switching problem.

Instead, however, we adopt a framework that is aimed at minimizing the Bayes risk relative to a Kullback–Leibler loss. The main idea is to find a configuration with distribution that is close to the corresponding 'true' distribution in terms of Bayes risk. To do this, we post-process the MCMC sample as follows.

(a) Find the positions of the actors that minimize the estimated Bayes risk among all positions.
(b) Procrustes transform the posterior draws of latent positions and, using the same transformation matrix, transform the cluster means and covariances.
(c) Find the cluster membership probabilities of the actors that minimize the estimated Bayes risk among all permutations of the cluster labels.

The general approach is due to Stephens (2000), and step (c) closely follows his solution to the label switching problem. The technical details of the steps are given in Appendix A.

4. Choosing the number of clusters

We recast the problem of choosing the number of clusters as one of model selection. Each number of clusters corresponds to a different statistical model, and we develop a Bayesian approach to comparing the resulting models.

One simple approach to model selection for the latent position cluster model is based on the two-stage maximum likelihood estimation method of Section 3.1. We first compute the maximum likelihood estimates of the latent positions by using the latent space model of Hoff *et al.* (2002). We then carry out model-based clustering of the resulting estimated latent positions, computing the Bayes information criterion BIC for each different number of groups, and choosing the number of groups with the highest values of BIC, as described by Dasgupta and Raftery (1998) and Fraley and Raftery (2002). As we shall see, however, this does not perform well, and instead we develop an approach that is based on the fully Bayesian estimation method of Section 3.2.

The standard Bayesian approach to model selection is to compute the posterior model probability of each of the competing models (Kass and Raftery, 1995). If we want to select a single model, we select the model with the highest posterior probability. The posterior model probability is proportional to the integrated likelihood for the model times the prior model probability. The integrated likelihood is obtained by integrating the likelihood times the prior over the model's parameter space. We assign equal prior probabilities to the models that we consider.

Here we use *conditional* posterior model probabilities, conditioning on an estimate of the latent positions, but integrating over the other parameters. We find the integrated likelihood of the observations *and* the estimated latent positions for each number of clusters. This was proposed by Oh and Raftery (2001, 2003) and worked well in a setting that was similar to the present one. There are several reasons for taking this approach. When selecting a model, we are

typically selecting an estimated configuration for visualization and interpretation, so it makes sense to evaluate the specific configuration of latent positions that will be used, rather than an average over the distribution of latent positions. When comparing different numbers of clusters, the dimension of the latent position parameter set that we condition on is the same regardless of the number of clusters. Finally, the dimension of the set of latent positions is high, and this can make it difficult to compute the integrated likelihood in a stable way.

For each value of the number of clusters, G, considered, we estimate the integrated likelihood of $(Y, \hat{Z}|G)$, with $\hat{Z}$ being a posterior estimate of the position of the actors. We choose the value of G that gives the largest value of $P(Y, \hat{Z}|G)$. Letting $\theta = (\mu_g, \lambda_g, \sigma_g^2)_{g=1}^G$, the integrated likelihood is

$$\begin{aligned} P(Y, \hat{Z}|G) &= \int P(Y, \hat{Z}|\beta, \theta)\ p(\beta)\ p(\theta)\ \mathrm{d}\beta\ \mathrm{d}\theta \\ &= \int P(Y|\hat{Z}, X, \beta)\ P(\hat{Z}|\theta)\ p(\beta)\ p(\theta)\ \mathrm{d}\beta\ \mathrm{d}\theta \\ &= \int P(Y|\hat{Z}, X, \beta)\ p(\beta)\ \mathrm{d}\beta \int P(\hat{Z}|\theta)\ p(\theta)\ \mathrm{d}\theta, \end{aligned}$$

where all terms are conditional on G.

The first integral on the right-hand side is the integrated likelihood for logistic regression of the observed ties conditional on the latent positions and the observed attributes, and the second integral is the integrated likelihood for the mixture model describing the latent positions. We approximate both of these integrals by using the Bayesian information criterion approximation (Schwarz, 1978). The BIC approximation for the integrated likelihood of a model for data D with n_{param} parameters θ and n_{obs} observations is

$$2\ \log\{P(D)\} \approx 2\ \log\{P(D|\hat{\theta})\} - n_{\mathrm{param}} \log(n_{\mathrm{obs}}). \tag{10}$$

The BIC approximation for the logistic regression is

$$\mathrm{BIC}_{\mathrm{lr}} = 2\ \log[P\{Y|\hat{Z}, \hat{\beta}(\hat{Z})\}] - d_{\mathrm{logit}}\ \log(n_{\mathrm{logit}}),$$

where $\hat{\beta}(\hat{Z})$ is the maximum likelihood estimator of β given that the latent positions are $Z = \hat{Z}$, $d_{\mathrm{logit}} = \dim(\beta)$ is the number of parameters in the logistic regression model and n_{logit} is the number of ties in the data. A possible alternative choice for n_{logit} is the number of possible ties, $n(n-1)$. We chose n_{logit} to be the number of actual rather than possible ties, on the basis of arguments that are analogous to those of Volinsky and Raftery (2000).

The BIC approximation for the mixture model is

$$\mathrm{BIC}_{\mathrm{mbc}} = 2\ \log[P\{\hat{Z}|\hat{\theta}(\hat{Z})\}] - d_{\mathrm{mbc}}\ \log(n),$$

where d_{mbc} is the number of parameters in the clustering model, and $\hat{\theta}(\hat{Z})$ is the maximum likelihood estimator of θ given that the latent positions are $Z = \hat{Z}$. Our final approximation is

$$\mathrm{BIC} = \mathrm{BIC}_{\mathrm{lr}} + \mathrm{BIC}_{\mathrm{mbc}}.$$

For both $\mathrm{BIC}_{\mathrm{lr}}$ and $\mathrm{BIC}_{\mathrm{mbc}}$, we use the minimum Kullback–Leibler estimates of the latent positions in the maximization of the likelihoods.

5. Examples

5.1. Example 1: liking between monks

We consider the social relations between 18 monks in an isolated American monastery (Samp-

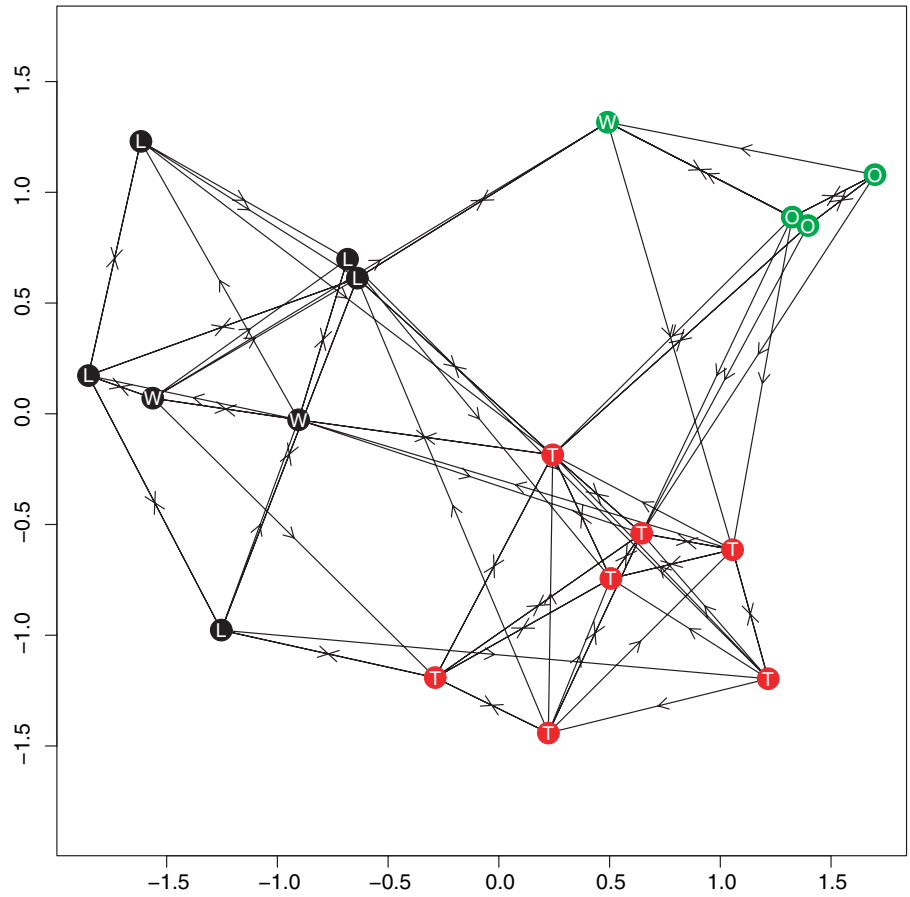

Fig. 1. Relationships between monks within a monastery: groups from two-stage maximum likelihood estimation of the latent position cluster model with three groups are shown by the colours of the nodes; a grouping given by Sampson (1969) is shown by the letters T (Turks), L (Loyal Opposition), O (Outcasts) and W (Waverers) (→, ties (i.e. the data))

son, 1969; White *et al.*, 1976). While resident at the monastery, Sampson collected extensive sociometric information by using interviews, experiments and observation. Here we focus on the social relation of 'liking'. We say that a monk has the social relation of 'like' to another monk if he ranked that monk in the top three monks for positive affection in any of three interviews given over a 12-month period.

We first consider the two-stage maximum likelihood estimation method, and the associated model selection approach. The maximum likelihood latent space positions from the first stage of the method are shown in Fig. 1. BIC from the model-based clustering of the second stage chose only one cluster. If we specify there to be three clusters, we obtain the estimated clusters that are shown in Fig. 1.

The data that were collected by Sampson (1969) have received much attention in the social networks literature (White *et al.*, 1976; Wasserman and Faust, 1994). Sampson provided a description of the clustering based on information that was collected at the end of the study period. He identified three main groups: the Young Turks (seven members), the Loyal Opposition (five members) and the Outcasts (three members). The other three monks wavered between the Loyal

Table 1. Two-stage maximum likelihood and Bayesian estimates of the parameters of the latent position cluster model for the relationship between monks within a monastery

Parameter	*Two-stage maximum likelihood*	*Lower 2.5%*	*Latent position cluster model posterior median*	*Upper 97.5%*	*Posterior standard deviation*	*Posterior median conditional on $\hat{Z}$*
β_0	3.475	1.028	1.820	2.830	0.458	
β_1	2.764	1.285	1.756	2.379	0.282	
μ_{11}	−1.188	−1.753	−1.376	−0.796	0.236	−1.081
μ_{12}	0.242	−0.407	0.072	0.503	0.235	0.141
μ_{21}	0.522	−0.170	0.305	0.814	0.250	0.420
μ_{22}	−0.849	−1.237	−0.772	−0.051	0.301	−0.571
μ_{31}	1.232	0.292	1.073	1.612	0.335	1.168
μ_{32}	1.032	0.018	0.686	1.156	0.286	0.756
σ_1	0.567	0.066	0.217	0.932	0.245	0.233
σ_2	0.446	0.044	0.137	0.696	0.186	0.152
σ_3	0.341	0.046	0.696	1.077	0.293	0.156
λ_1	0.389	0.222	0.389	0.500	0.060	0.389
λ_2	0.389	0.222	0.389	0.500	0.071	0.389
λ_3	0.222	0.167	0.222	0.444	0.068	0.222

Opposition and the Young Turks, which he described as being in intense conflict (Sampson (1969), page 370, and White *et al.* (1976), pages 752–753). The groups that were identified by Sampson are indicated by letters in Fig. 1. The data that we model here include only one of the relationships that Sampson considered in his analysis.

In our two-stage solution, the Young Turks form their own group, and the Loyal Opposition and Outcasts are each contained in separate groups. The Waverers are split, with one clustered with the Outcasts and the other two with the Loyal Opposition. White *et al.* (1976) developed block models for social relations within the monastery based on eight positive and negative social relations. Although their methodology was different, their primary objective was clustering of the monks. Their model found three groups in the monastery; the groups coincide exactly with those from our two-stage method when the number of groups is constrained to be 3 (White *et al.* (1976), page 753). Our model yields the same results as theirs, even though they used much more information.

We then fitted our Bayesian model using MCMC sampling with 5000 burn-in iterations that were discarded, and a further 30000 iterations, of which we kept every 30th value. Visual display of trace plots and more formal assessments of convergence (e.g. Raftery and Lewis (1996)) indicated that this gave results that were sufficiently accurate for our purposes. The parameter estimates from the two-stage and Bayesian methods are shown in Table 1.

The plot of the BIC-values is given in Fig. 2 and indicates a clear choice of three clusters. This is in line with the previous research.

Fig. 3 shows the minimum Kullback–Leibler estimates of the social positions of the monks for the three-cluster model. The monks are well separated into the three clusters—even the monk from the Loyal Opposition who had five ties to the other monks within his group and three ties to the Young Turks is now well separated from the Young Turks. The Young Turks are also more tightly clustered than the Loyal Opposition. Sampson's analysis indicated larger heterogeneity of actors within the Loyal Opposition group. This is reflected in the fissure between two components of the Loyal Opposition. The Outcasts are also closely bound, and the Waverer who is

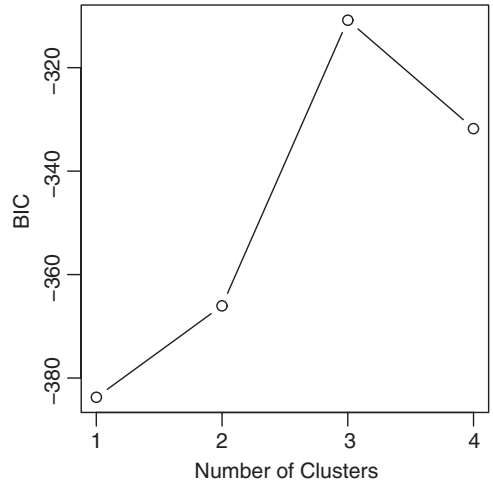

Fig. 2. BIC-plot for the latent position clustering model of the relationship between monks within a monastery

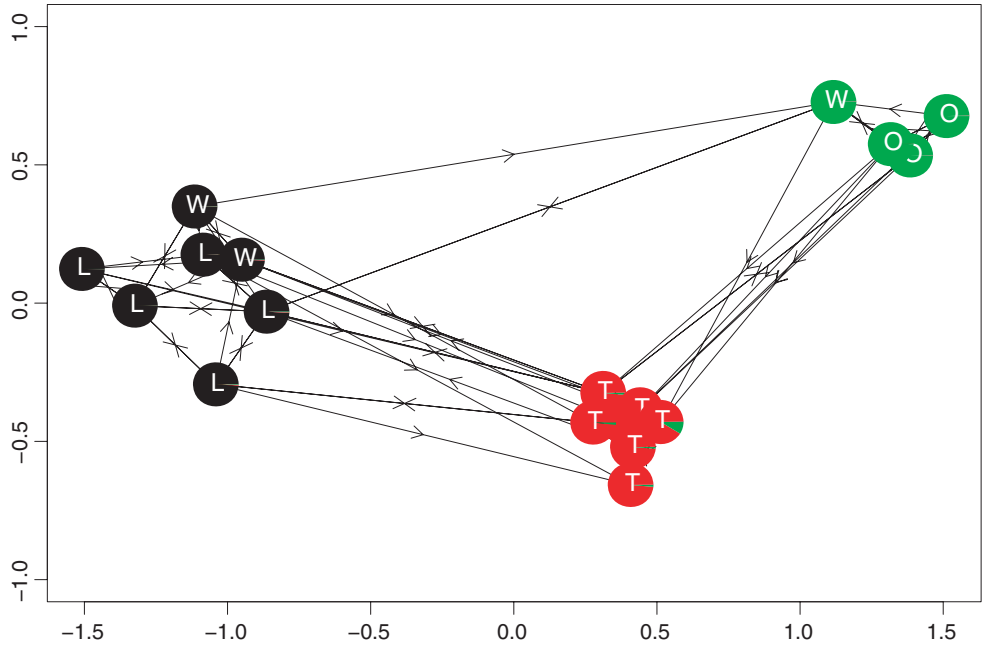

Fig. 3. Estimates of clusters and latent positions for the relationship between monks within a monastery from the Bayesian estimation of the latent position cluster model: the probability of assignment to each latent cluster is shown by a pie chart

clustered with them is the farthest from the others. Overall the Bayesian estimate of the latent position cluster model produces greater distinctions between the groups than the two-stage estimate and firmly identifies the grouping of the Waverers.

The uncertainty in the cluster assignments is shown in Fig. 3, where the cluster assignment probabilities for each actor are shown as pie charts. We see that most actors have almost no probability of belonging to any other cluster—except for one of the Young Turks.

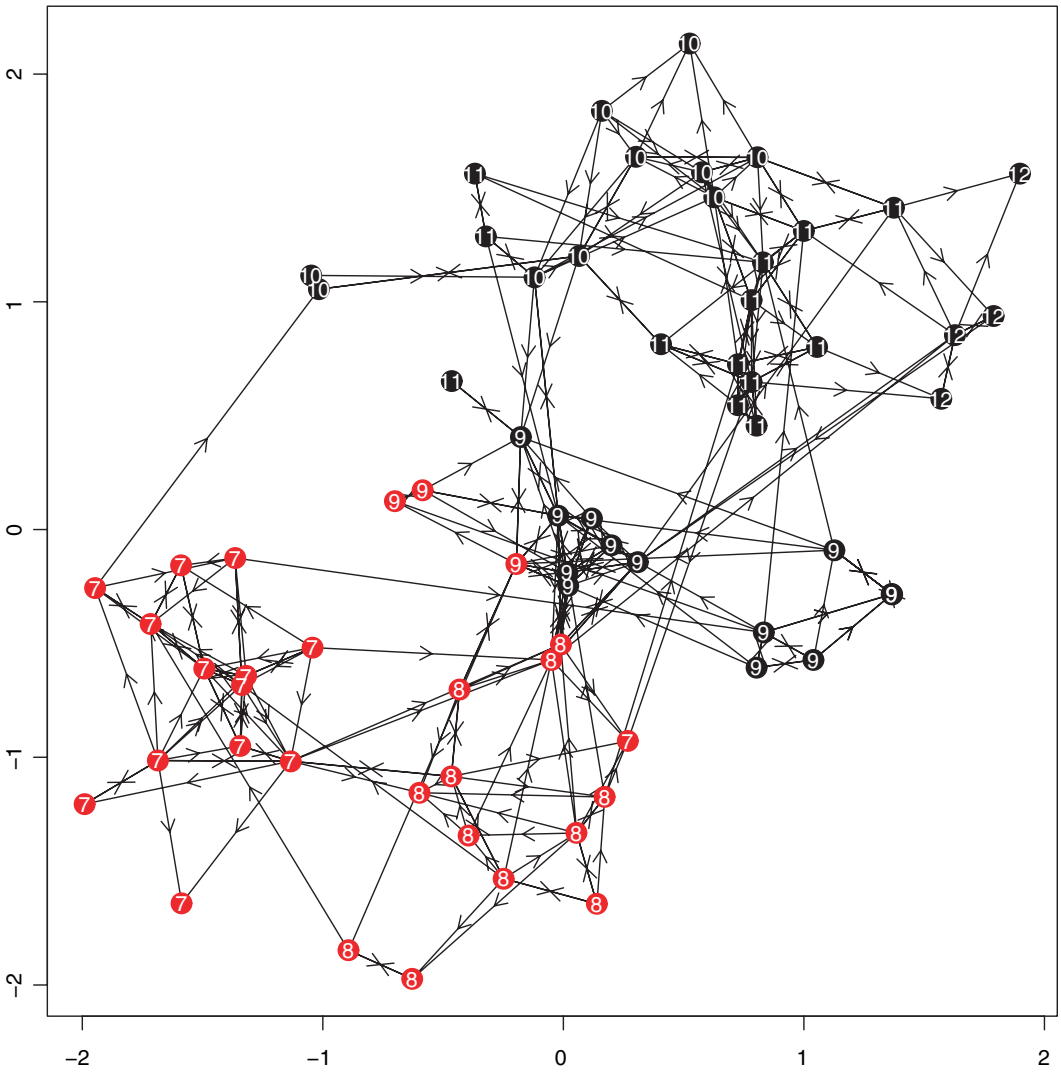

Fig. 4. Two-stage maximum likelihood estimates of the latent positions and clusters for the adolescent health data, where the number of clusters (2) is chosen by BIC: clusters are shown by colour with actual grades shown as numbers

The two-stage method performs well here when the number of clusters is known in advance. However, the Bayesian method correctly estimates the number of groups and also yields tighter estimates of the latent positions. This is because it borrows strength from the clustering information when estimating the latent positions. The Bayesian approach allows the uncertainty in cluster assignment to take into account the uncertainty in actor position and vice versa, and this turns out to be important for these data.

5.2. Example 2: adolescent health

The second social network is from the National Longitudinal Study of Adolescent Health. The study is a school-based longitudinal study of the health-related behaviours of adolescents and

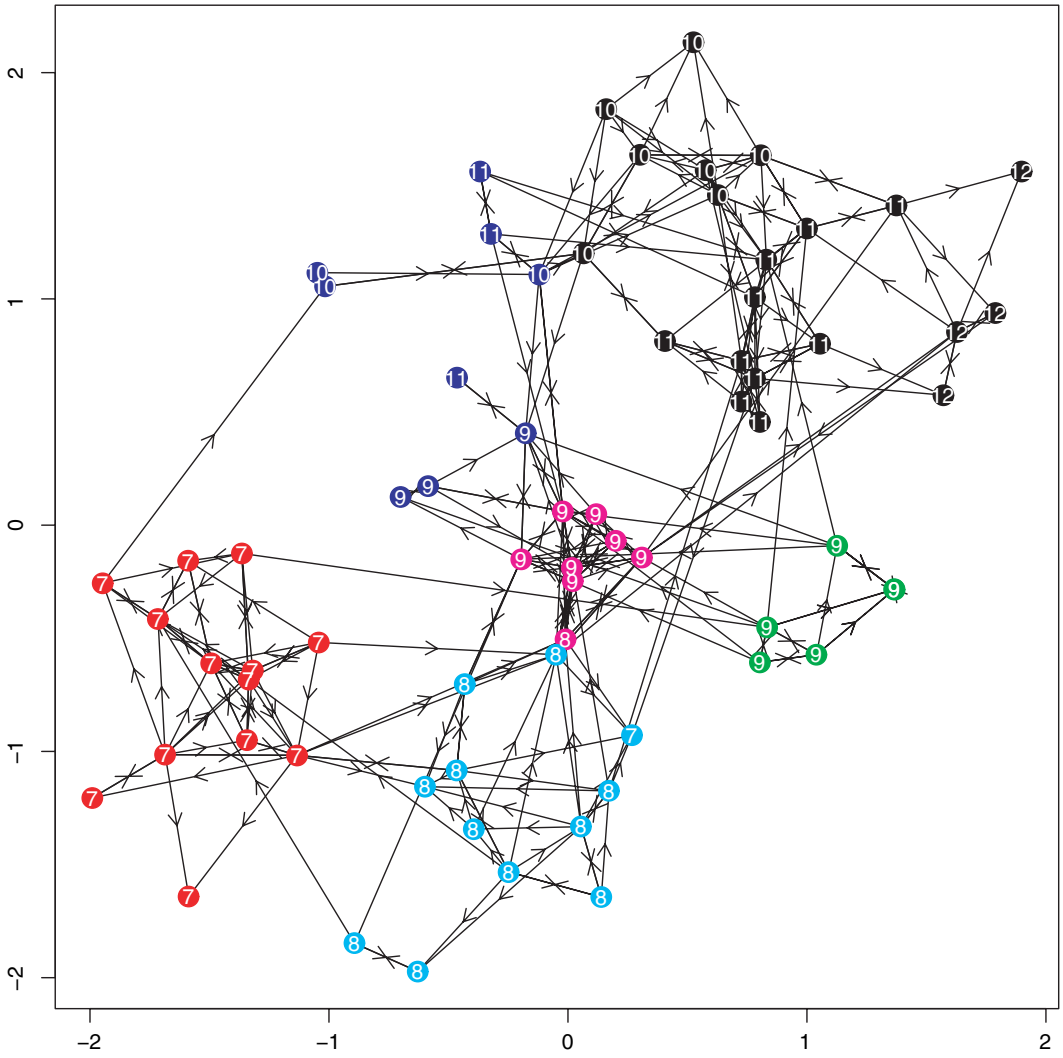

Fig. 5. Clusters from two-stage maximum likelihood estimates of the latent position cluster model for the adolescent health data, where the number of clusters is constrained to be 6: clusters are shown by colour with actual grades shown as numbers; there are six green points, representing students from grade 9, two of which are coincident

their outcomes in young adulthood. The study design sampled 80 high schools and 52 middle schools from the USA that were representative with respect to region of the country, urbanicity, school size, school type and ethnicity (Harris *et al.*, 2003). In 1994–1995 an in-school questionnaire was administered to a nationally representative sample of students in grades 7–12. In addition to demographic and contextual information, each respondent was asked to nominate up to five boys and five girls within the school whom they regarded as their best friends. Thus each student could nominate up to 10 students within the school (Udry, 2003).

Here we consider a single school of 71 adolescents from grades 7–12. We consider the friendship nominations between those who have either nominated at least one other adolescent as their friend or who have been nominated at least once as the friend of another adolescent. Two

Table 2. Adolescent health data: clusters from two-stage maximum likelihood estimation of the latent position cluster model with six clusters, compared with the student's grades

Grade	*Results for the following clusters:*					
	1	*2*	*3*	*4*	*5*	*6*
7	13	1	0	0	0	0
8	0	11	1	0	0	0
9	0	0	7	6	3	0
10	0	0	0	0	3	7
11	0	0	0	0	3	10
12	0	0	0	0	0	4

adolescents who had no ties in the network were excluded. The remaining 69 adolescents form a connected directed network with nodes the adolescents and nominations the ties.

We fitted the latent position cluster model without using the grade of the adolescents. Instead we used the grade information for assessing the clustering. The two-stage maximum likelihood estimates of the latent positions are given in Fig. 4. The approximate BIC-values based on the two-stage maximum likelihood estimates chose two clusters, which seems a poor choice given the grade information. When we required six clusters, we obtained the results that are shown in Fig. 5. Now the clusters have a loose correspondence to grade and most actors of the same grade are close to each other.

The correspondence between clusters and grade is shown in Table 2. The seventh- and eighth-grade adolescents belong to two clusters that are mostly homogeneous with respect to grade. The ninth-grade adolescents fall into two clusters. The 10th-, 11th- and 12th-grade adolescents fall into two clusters, one of which includes all four 12th graders.

We fitted our Bayesian model using MCMC sampling with 50 000 burn-in iterations that were discarded, and a further 2 million iterations, of which we kept every 1000th value. The result-

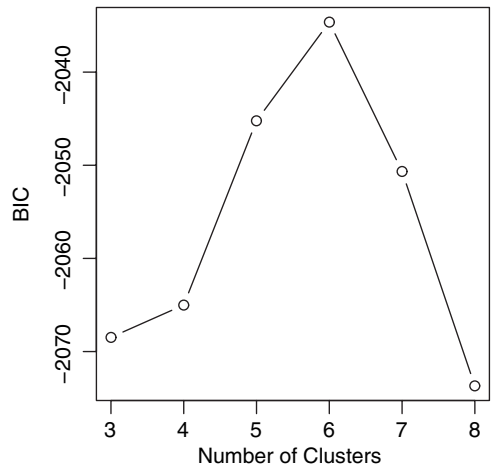

Fig. 6. BIC-plot for the latent position cluster model of the adolescent health network

Table 3. Two-stage maximum likelihood estimates and Bayesian estimates of the parameters of the latent position cluster model for the adolescent health network

Parameter	*Latent position mixture model estimate*	*Lower 2.5%*	*Latent position cluster model posterior median*	*Upper 97.5%*	*Posterior standard deviation*	*Posterior median conditional on $\hat{Z}$*
β_0	1.031	1.001	1.394	1.812	0.208	
β_1	5.205	3.464	3.962	4.549	0.276	
μ_{11}	−1.495	−1.579	−1.168	−0.485	0.271	−1.126
μ_{12}	−0.704	−1.493	−0.532	0.091	0.405	−0.610
μ_{21}	−0.278	−1.130	−0.348	0.385	0.378	−0.241
μ_{22}	−1.276	−1.574	−1.048	0.177	0.448	−1.163
μ_{31}	0.051	−0.484	0.196	0.758	0.304	0.180
μ_{32}	−0.165	−0.819	−0.179	0.526	0.315	−0.224
μ_{41}	1.088	0.008	0.999	1.546	0.391	1.028
μ_{42}	−0.384	−0.940	−0.276	0.594	0.385	−0.392
μ_{51}	−0.497	−0.430	0.349	1.019	0.364	0.599
μ_{52}	0.808	−0.456	1.243	1.683	0.317	1.118
μ_{61}	0.837	−0.863	−0.057	1.052	0.490	−0.172
μ_{62}	1.139	−0.310	0.830	1.476	0.464	0.880
σ_1	0.368	0.169	0.368	0.895	0.216	0.231
σ_2	0.402	0.192	0.495	1.012	0.267	0.257
σ_3	0.166	0.123	0.291	1.035	0.325	0.107
σ_4	0.204	0.178	0.456	1.022	0.294	0.207
σ_5	0.450	0.262	0.538	0.985	0.448	0.257
σ_6	0.514	0.152	0.479	1.064	0.323	0.233
λ_1	0.188	0.058	0.159	0.348	0.070	0.188
λ_2	0.174	0.043	0.159	0.348	0.075	0.159
λ_3	0.116	0.029	0.116	0.333	0.072	0.130
λ_4	0.087	0.043	0.130	0.319	0.070	0.145
λ_5	0.130	0.058	0.217	0.406	0.094	0.188
λ_6	0.304	0.029	0.145	0.348	0.085	0.188

ing BIC (Fig. 6) chose six clusters. This is the same as the number of grades, and the clusters correspond roughly to the grades, so the BIC-estimate has some face validity.

The parameter estimates for both the two-stage and the Bayesian estimates of the latent position cluster model are shown in Table 3. Fig. 7 shows the Bayesian estimates of the children's social positions for the six-cluster model.

As we would expect the students tend to be linked to others in their own grade; we can compare the clusters that are identified by the model with the grades. As the model is unaware of the grades of the students, we are asking the model to identify a latent clustering that should be a partial surrogate for grade. The clusters correspond quite well to grades. This comparison is summarized in Table 4.

The seventh-grade adolescents are in their own well-separated cluster, with the exception of one seventh grader whose only friends are eighth graders (possibly a student who had been held back). The eighth graders are mostly in their own cluster, with two having stronger ties to the ninth-grade class and so being incorporated in that cluster. The ninth-grade class is split into two clusters, the social magentas and the cliquey greens. The magenta ninth-grade cluster has many ties to other clusters, whereas the ties from the green ninth-grade cluster to other clusters are mostly to the other (magenta) ninth-grade cluster. The 11th grader whose only friend is a green ninth grader is more likely to belong to the green cluster ninth grade than to any other.

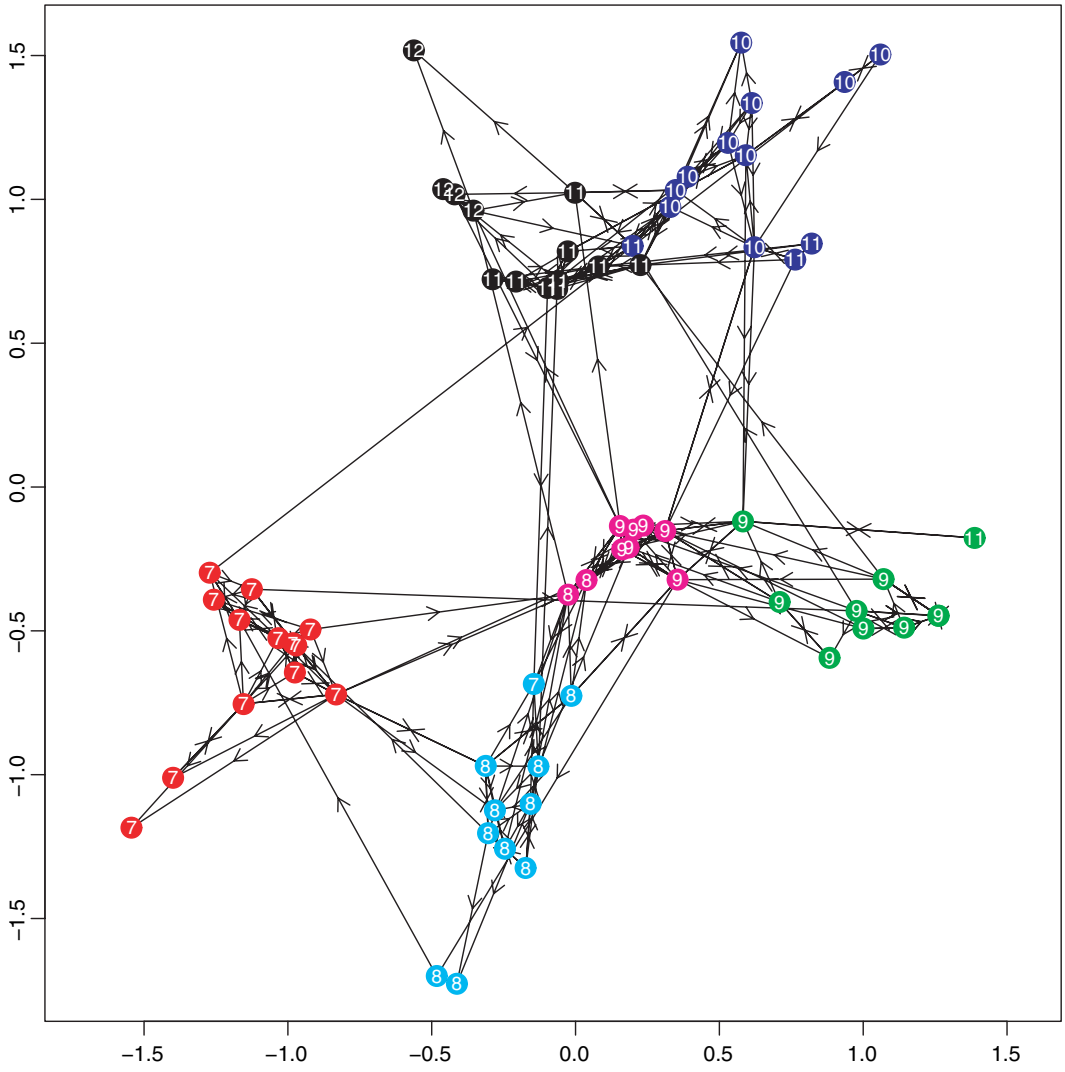

Fig. 7. Bayesian estimates of posterior clusters and latent positions for the friendship network in the adolescent health school: latent clusters are shown by colour with actual grades shown as numbers

The 10th-, 11th- and 12th-grade classes belong to two clusters which are very close in the latent social space. The 10th-grade class is entirely contained within the blue cluster, and most of the 11th graders and all the 12th graders are in the black cluster.

Thus the model has identified the strong tendency of students to form ties with others in their own grade, as the clusters line up well but not perfectly with the grades. There is also a more subtle tendency for the within-grade cohesion to weaken as students move up in the school, from the tightly linked seventh graders to the more loosely tied students in the top three grades, who associate more easily with students in grades other than their own. This may reflect a tendency of students to form links increasingly based on common interests and personal affinity and less on the grade that they happen to be in, as they gain seniority.

Fig. 8 shows the cluster assignment probabilities for each student. Students in the magenta cluster also have a significant probability of belonging to the cyan eighth grade cluster, whereas

Table 4. Clusters from the latent position cluster model compared with the student's grades for the adolescent health network†

Grade	*Results for the following clusters:*					
	1	*2*	*3*	*4*	*5*	*6*
7	13	1	0	0	0	0
8	0	10	2	0	0	0
9	0	0	7	9	0	0
10	0	0	0	0	10	0
11	0	0	0	1	3	9
12	0	0	0	0	0	4

†Note the concordance between the clusters and the actual grades.

the magenta ninth graders have significant probability of belonging only to the magenta and green clusters. The uncertainty in cluster assignment between the blue and black clusters is clearly visible.

The Bayesian method provides a much better estimate of the number of groups than the two-stage maximum likelihood estimation approach. The clusters are well defined in terms of both their positions in space and their correspondence to the grades. This is reflected in the estimates of the positions and the uncertainty in cluster membership (Fig. 8).

6. Discussion

We have proposed a new model for social networks: the latent position cluster model. This captures three important commonly observed features of networks, namely transitivity, homophily on attributes and clustering. We have developed two methods for estimating the latent positions and the model parameters: a simple two-stage maximum likelihood procedure and a fully Bayesian approach using MCMC sampling. We have also developed a Bayesian approach to finding the number of clusters in the data. The methods work well for two data sets. The two-stage maximum likelihood estimation approach works fairly well and is simple to implement, whereas the fully Bayesian approach performs better but is more complex.

The model can be thought of as a restriction of the latent position model to represent coherent groups of actors within the network better. It links the observed patterns of ties to latent positions where the latter may be partially determined by unobserved attributes of the actors in addition to social forces.

Our approach could be extended in several ways. We have developed it as a model for directed ties, but it could easily be adapted to data involving undirected ties: instead of a likelihood component for each ordered pair of actors (i, j) and (j, i), there would then be only one, for the unordered pair (i, j). We have specified our model as a model for binary ties: present or absent. However, network ties often have non-binary values, such as counts (e.g. the number of phone calls between two people), or continuous values (e.g. the volume of trade between two countries). Our model can be easily extended to these situations by replacing the binary logistic regression of equation (2) by a generalized linear model or another specification of dependence.

We have required the dimension of the latent social space to be specified by the user. It could be desirable to estimate this from the data, and this is possible by using methods that are similar

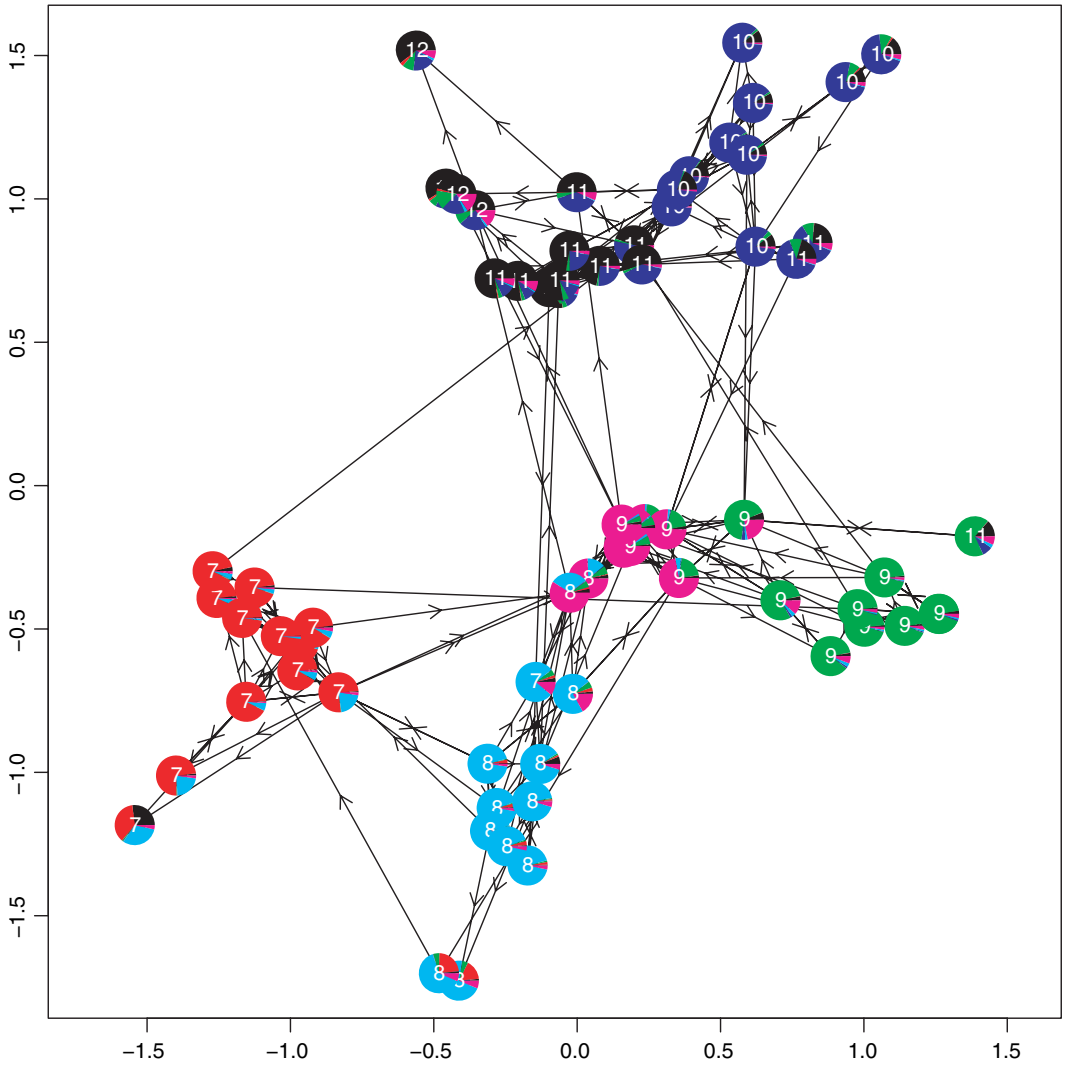

Fig. 8. Pie charts for posterior probabilities of cluster assignment for each actor, at the Bayesian estimates of posterior latent positions for the friendship network in the adolescent health school: the students' grades are shown as numbers

to those in Section 4, as developed in a slightly different context by Oh and Raftery (2001). The use of two dimensions leads to easy visualization, but higher dimensions may be needed to represent the network adequately, especially for larger networks.

Two important local characteristics of networks are the tendency for a tie within a dyad to be reciprocated, and for triads to be transitive. The latent position cluster model determines these characteristics on the basis of the distances between the actors and their covariate values. The propensity for reciprocity and transitivity in a data set may be higher or lower than that prescribed by the model. For the applications that were considered in this paper the posterior mean reciprocity and transitivity are consistent with the levels in the data. However, the model may need to be extended to model reciprocity and transitivity in other networks explicitly.

One important aspect of social networks that our model does not explicitly incorporate is the differing tendency of actors to send and receive ties. The model could be extended to this situation by including random effects for the propensity of actors to send and receive ties in equation (2), similarly to van Duijn *et al.* (2004) and Hoff (2005). In our examples, the propensities of the actors to receive ties (although not to send them) differed considerably, and our model reflects this sufficiently well, but if the differences were much more extreme the model as currently specified might have difficulties.

One use of social network models is to provide inputs to models of larger systems of which the networks are part. An important example of this is epidemiological modelling of the spread of contagious diseases (Kretzschmar and Morris, 1996; Bearman *et al.*, 2004; Eames and Keeling, 2004; Eubank *et al.*, 2004). It is easy to simulate realizations from our model conditional on estimated or specified parameters and, by using draws from the posterior distribution, one can simulate a realistic range of scenarios. Often there is interest in simulating an entire population for which network data are available for only a small part. This could be done using our model, if necessary by combining it with a simple model such as a Poisson process for the means of clusters that were not represented in the data that were analysed. Although feasible, our method is computationally demanding, and so for larger networks more computationally efficient versions of our estimation methods should be sought.

The model has many potential application areas. The applications in this paper focus on social relations where the ties represents positive affect. Network phenomena are ubiquitous in the sciences (e.g. biology, information sciences and epidemiology). The model applies to diverse types of relationships (e.g. biological interaction, exchange, co-citation, common affiliations or food source) and nodes (e.g. proteins, villages, authors and organizations, or animals).

An R package called `latentnet` implementing the procedures in this paper is publicly available on the 'Comprehensive R archive network', at `http://cran.r-project.org`.

Acknowledgements

The authors are listed in alphabetical order. The research of Tantrum and Handcock was supported by NIDA grant DA012831 and NICHD grant HD041877. Raftery's research was supported by National Institutes of Health grant 8 R01EB 002137-02. The work was completed while Tantrum held a post-doctoral position funded by the National Institutes of Health at the Center for Studies in Demography and Ecology at the University of Washington. The authors are grateful to Peter Hoff and four reviewers for very helpful comments.

Appendix A: Identifiability of positions and cluster labels

Here we give some details of the steps in the algorithm to post-process the MCMC output for identifying the positions of the actors, the cluster means and variances, and the cluster membership probabilities.

A.1. Actor positions via minimum Kullback–Leibler divergence

Let $\mathrm{KL}(Z, \beta; \tilde{Z}, \tilde{\beta})$ denote the Kullback–Leibler divergence of the distribution of Y at Z and β to the distribution at $\tilde{Z}$ and $\tilde{\beta}$. Let $\eta(Z, X, \beta) = [\eta_{ij}(Z, X, \beta)]$, where $\eta_{ij}(Z, X, \beta)$ is the log-odds of a tie given by the right-hand side of equation (2). Equation (1) can be re-expressed as

$$P(Y|Z, X, \beta) = \frac{\exp\{\eta^{\mathrm{T}}(Z, X, \beta)y\}}{c(Z, X, \beta)}$$

where $\eta(Z, X, \beta)$ is vectorized in canonical order. The Kullback–Leibler divergence is

$$\mathrm{KL}(Z,\beta;\tilde{Z},\tilde{\beta}) = \sum_{y} \log\left\{\frac{P(Y|Z,X,\beta)}{P(Y|\tilde{Z},X,\tilde{\beta})}\right\} P(Y|Z,X,\beta)$$
$$= (\eta(Z,X,\beta) - \eta(\tilde{Z},X,\tilde{\beta}))^{\mathrm{T}}\, E_{Z,X,\beta}[Y] - \log\left\{\frac{c(\tilde{Z},X,\tilde{\beta})}{c(Z,X,\beta)}\right\},$$

where the sum is over all possible values of y and

$$E_{Z,X,\beta}[Y]_{ij} = \frac{\exp\{\eta_{ij}(Z,X,\beta)\}}{1+\exp\{\eta_{ij}(Z,X,\beta)\}}.$$

We seek the values of $\tilde{Z}$ and $\tilde{\beta}$ that minimize the loss function $\mathrm{KL}(Z,\beta;\tilde{Z},\tilde{\beta})$. As the true values $\tilde{Z}$ and $\tilde{\beta}$ are unknown we focus on values of $\tilde{Z}$ and $\tilde{\beta}$ that minimize the corresponding Bayes risk, i.e. the posterior expected loss:

$$E_{Z,\beta|Y_{\mathrm{obs}}}[\mathrm{KL}(Z,\beta;\tilde{Z},\tilde{\beta})]$$
$$= E_{Z,\beta|Y_{\mathrm{obs}}}[\eta(Z,X,\beta)^{\mathrm{T}}\, E_{Z,X,\beta}[Y]] - \eta(\tilde{Z},X,\tilde{\beta})^{\mathrm{T}}\, E[Y|Y_{\mathrm{obs}}] + E_{Z,\beta|Y_{\mathrm{obs}}}[\log\{c(Z,X,\beta)\}]$$
$$- \log\{c(\tilde{Z},X,\tilde{\beta})\},$$

where

$$E[Y|Y_{\mathrm{obs}}] = E_{Z,\beta|Y_{\mathrm{obs}}}[E_{Z,X,\beta}[Y]]$$

is the posterior mean of Y. As the first and third terms do not involve $\tilde{Z}$ or $\tilde{\beta}$ this is equivalent to maximizing

$$\frac{\exp\{\eta^{\mathrm{T}}(\tilde{Z},X,\tilde{\beta})\, E[Y|Y_{\mathrm{obs}}]\}}{c(\tilde{Z},X,\tilde{\beta})},$$

which can be done by using the likelihood maximization method that was previously described. In our procedure the posterior mean $E[Y|Y_{\mathrm{obs}}]$ is estimated from the MCMC sample, and so $\tilde{Z}$ and $\tilde{\beta}$ minimize the corresponding estimate of the Bayes risk.

A.2. Label switching via minimum Kullback–Leibler divergence

The idea of minimizing a Kullback–Leibler divergence to solve the label switching problem was introduced by Stephens (2000), and here we adapt his algorithm to our model.

Let $P(\theta) = [p_{ig}(\theta)]$, where $p_{ig}(\theta)$ denotes the probability of classifying actor i into cluster g given by equation (9). We write θ for the vector of parameter values $\{\lambda_g, \mu_g, \sigma_g^2\}_{g=1}^{G}$. To express uncertainty in the cluster memberships we use $Q = [q_{ig}]$, where q_{ig} denotes the probability that actor i is assigned to cluster g. For a given parameter vector θ, denote the Kullback–Leibler distance from the distribution $P(\theta)$ to the distribution Q by

$$\mathrm{KL}(\theta;Q) = \sum_{i,g} p_{ig}(\theta)\, \log\left\{\frac{p_{ig}(\theta)}{q_{ig}}\right\}.$$

Following Stephens (2000), we seek Q that minimizes the divergence over all permutations of the cluster labels. Specifically, let π be a permutation of $1,\ldots,g$ and $\pi(\theta) = \{\lambda_{\pi(1)},\ldots,\lambda_{\pi(g)}, \mu_{\pi(1)},\ldots,\mu_{\pi(g)}, \sigma^2_{\pi(1)},\ldots,\sigma^2_{\pi(g)}\}$ be the corresponding permutation of θ. Then we seek Q to minimize the loss function:

$$\min_{\pi\in\Upsilon}[\mathrm{KL}\{\pi(\theta);Q\}]$$

where Υ is the set of all permutations of $1,\ldots,g$. As the true value of θ is unknown we focus on values of Q that minimize the corresponding Bayes risk, i.e. the posterior expected loss. In our algorithm the Bayes risk is approximated by the mean loss over the MCMC sample, and Q is chosen to minimize this approximation. See Stephens (2000), algorithm 2, for the explicit computational steps.

References

Anderson, C. J. and Wasserman, S. S. (1987) Stochastic a posteriori blockmodels: construction and assessment. *Socl Netwrks*, **9**, 1–36.

Banfield, J. D. and Raftery, A. E. (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics*, **49**, 803–821.
Bearman, P. S., Moody, J. and Stovel, K. (2004) Chains of affection: the structure of adolescent romantic and sexual networks. *Am. J. Sociol.*, **110**, 44–91.
Breiger, R. L., Boorman, S. A. and Arabie, P. (1975) An algorithm for clustering relational data with application to social network analysis and comparison with multidimensional scaling. *J. Math. Psychol.*, **12**, 328–383.
Celeux, G., Hurn, M. and Robert, C. (2000) Computational and inferential difficulties with mixture posterior distribution. *J. Am. Statist. Ass.*, **95**, 957–970.
Dasgupta, A. and Raftery, A. E. (1998) Detecting features in spatial point processes with clutter via model-based clustering. *J. Am. Statist. Ass.*, **93**, 294–302.
Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.
Diebolt, J. and Robert, C. P. (1994) Estimation of finite mixture distributions through Bayesian sampling. *J. R. Statist. Soc.* B, **56**, 363–375.
Doreian, P., Batagelj, V. and Ferligoj, A. (2005) *Generalized Blockmodeling*. Cambridge: Cambridge University Press.
van Duijn, M. A. J., Snijders, T. A. B. and Zijlstra, B. H. (2004) p_2: a random effects model with covariates for directed graphs. *Statist. Neerland.*, **58**, 234–254.
Eames, K. T. D. and Keeling, M. J. (2004) Monogamous networks and the spread of sexually transmitted diseases. *Math. Biosci.*, **189**, 115–130.
Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z. and Wang, N. (2004) Modelling disease outbreaks in realistic urban social networks. *Nature*, **429**, 180–184.
Faust, K. (1988) Comparison of methods for positional analysis: structural and general equivalence. *Socl Netwrks*, **10**, 313–341.
Fienberg, S. E. and Wasserman, S. S. (1981) Categorical data analysis of single sociometric relations. *Sociol. Methodol.*, **11**, 156–192.
Fraley, C. and Raftery, A. E. (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput. J.*, **41**, 578–588.
Fraley, C. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis and density estimation. *J. Am. Statist. Ass.*, **97**, 611–631.
Fraley, C. and Raftery, A. E. (2003) Enhanced model-based clustering, density estimation and discriminant analysis software: MCLUST. *J. Classificn*, **20**, 263–286.
Frank, O. and Strauss, D. (1986) Markov graphs. *J. Am. Statist. Ass.*, **81**, 832–842.
Freeman, L. C. (1996) Some antecedents of social network analysis. *Connections*, **19**, 39–42.
Harris, K. M., Florey, F., Tabor, J., Bearman, P. S., Jones, J. and Udry, R. J. (2003) The national longitudinal of adolescent health: research design. *Technical Report*. Carolina Population Center, University of North Carolina, Chapel Hill. (Available from `http://www.cpc.unc.edu/projects/addhealth/design`.)
Hoff, P. D. (2005) Bilinear mixed-effects models for dyadic data. *J. Am. Statist. Ass.*, **100**, 286–295.
Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002) Latent space approaches to social network analysis. *J. Am. Statist. Ass.*, **97**, 1090–1098.
Hoff, P. D. and Ward, M. D. (2004) Modeling dependencies in international relations networks. *Polit. Anal.*, **12**, 160–175.
Holland, P. W. and Leinhardt, S. (1981) An exponential family of probability distributions for directed graphs (with discussion). *J. Am. Statist. Ass.*, **76**, 33–65.
Kass, R. and Raftery, A. E. (1995) Bayes factors. *J. Am. Statist. Ass.*, **90**, 773–795.
Kretzschmar, M. and Morris, M. (1996) Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.*, **133**, 165–195.
Lazarsfeld, P. and Merton, R. (1954) Friendship as social process: a substantive and methodological analysis. In *Freedom and Control in Modern Society* (eds M. Berger, T. Abel and C. Page), pp. 18–66. New York: Van Nostrand.
Liotta, G. (ed.) (2004) Graph drawing. *Lect. Notes Comput. Sci.*, **2912**.
Lorrain, F. and White, H. (1971) Structural equivalence of individuals in social networks blockstructures with covariates. *J. Math. Sociol.*, **1**, 49–80.
McFarland, D. D. and Brown, D. J. (1973) Social distance as a metric: a systematic introduction to smallest space analysis. In *Bonds of Pluralism: the Form and Substance of Urban Social Networks* (ed. E. O. Laumann), pp. 213–253. New York: Wiley.
McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001) Birds of a feather: homophily in social networks. *A. Rev. Sociol.*, **27**, 415–444.
Newman, M. E. J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.
Nowicki, K. and Snijders, T. A. B. (2001) Estimation and prediction for stochastic blockstructures. *J. Am. Statist. Ass.*, **96**, 1077–1087.
Oh, M. S. and Raftery, A. E. (2001) Bayesian multidimensional scaling and choice of dimension. *J. Am. Statist. Ass.*, **96**, 1031–1044.

Oh, M. S. and Raftery, A. E. (2003) Model-based clustering with dissimilarities: a Bayesian approach. *Technical Report 441*. Department of Statistics, University of Washington, Seattle.
Raftery, A. E. and Lewis, S. M. (1996) Implementing MCMC. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, D. J. Spiegelhalter and S. Richardson), pp. 115–130. London: Chapman and Hall.
Sampson, S. F. (1969) Crisis in a cloister. *PhD Thesis*. Cornell University, Ithaca.
Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.
Schweinberger, M. and Snijders, T. A. B. (2003) Settings in social networks: a measurement model. *Sociol. Methodol.*, **33**, 307–341.
Sibson, R. (1979) Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling. *J. R. Statist. Soc.* B, **41**, 217–229.
Snijders, T. (1991) Enumeration and simulation methods for 0-1 matrices with given marginals. *Psychometrika*, **56**, 397–417.
Snijders, T. A. B. and Nowicki, K. (1997) Estimation and prediction for stochastic block-structures for graphs with latent block structure. *J. Classificn*, **14**, 75–100.
Snijders, T. A. B., Pattison, P., Robins, G. L. and Handcock, M. S. (2005) New specifications for exponential random graph models. *Sociol. Methodol.*, **35**, in the press.
Stephens, M. (2000) Dealing with label switching in mixture models. *J. R. Statist. Soc.* B, **62**, 795–809.
Tallberg, C. (2005) A Bayesian approach to modeling stochastic blockstructures with covariates. *J. Math. Sociol.*, **29**, 1–23.
Udry, R. J. (2003) The national longitudinal of adolescent health: (add health), waves i and ii, 1994-1996; wave iii, 2001-2002. *Technical Report*. Carolina Population Center, University of North Carolina, Chapel Hill.
Volinsky, C. T. and Raftery, A. E. (2000) Bayesian information criterion for censored survival models. *Biometrics*, **56**, 256–262.
Wasserman, S. S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
White, H. C., Boorman, S. A. and Breiger, R. L. (1976) Social-structure from multiple networks: I, Blockmodels of roles and positions. *Am. J. Sociol.*, **81**, 730–780.