J. R. Statist. Soc. A (2007)
170, *Part* 2, *pp.* 301–354

Model-based clustering for social networks

Mark S. Handcock and Adrian E. Raftery

University of Washington, Seattle, USA

and Jeremy M. Tantrum

Microsoft adCenter Labs, Redmond, 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

 0964–1998/07/170301

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.*, 2006).

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 (*blockmodels*). 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), Faust (1988) 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 blockmodel 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 $Z_i^* \sim \text{MVN}_d(Z_{it}, \delta_Z^2 I_d)$.
(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 the 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

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

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 BIC 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

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

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-

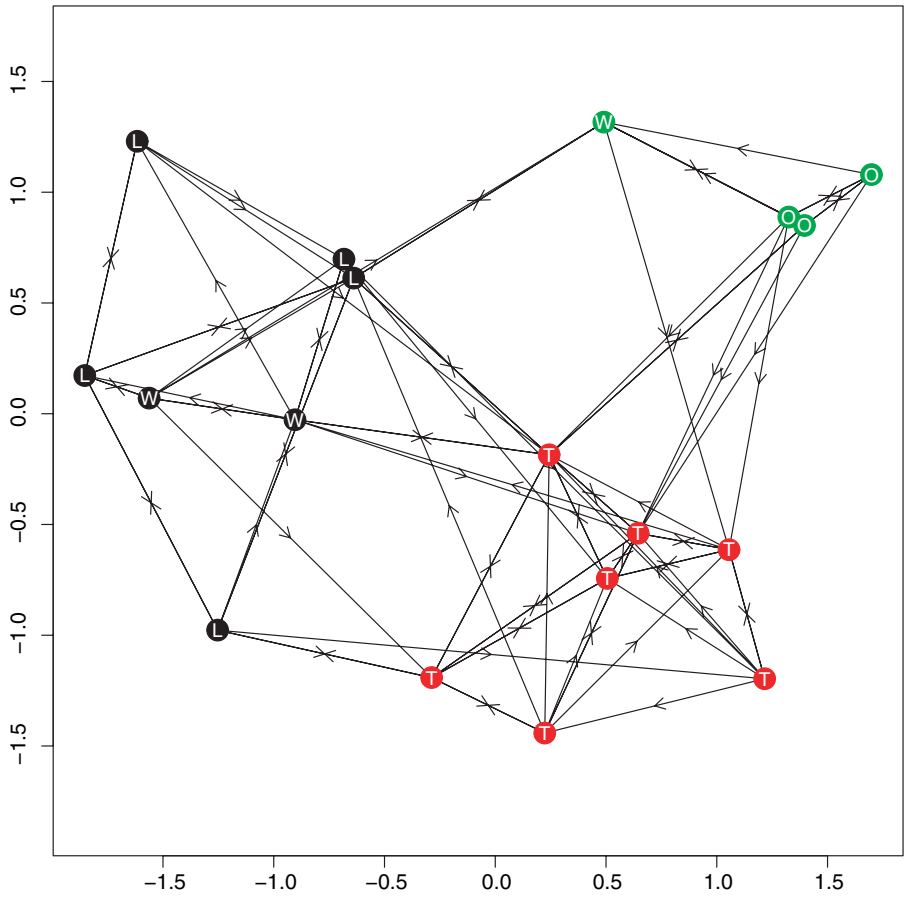

Fig. 1. Relationships between monks within a monastery: groups from two-stage maximum likelihood estimation of the latent position cluster model with three groups are shown by the colours of the nodes; a grouping given by Sampson (1969) is shown by the letters T (Turks), L (Loyal Opposition), O (Outcasts)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

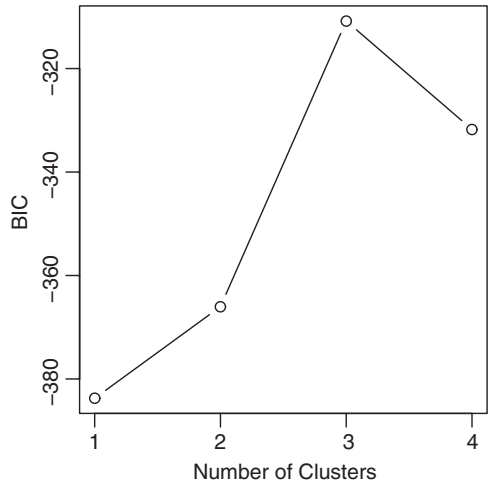

Fig. 2. BIC plot for the latent position clustering model of the relationship between monks within a monastery

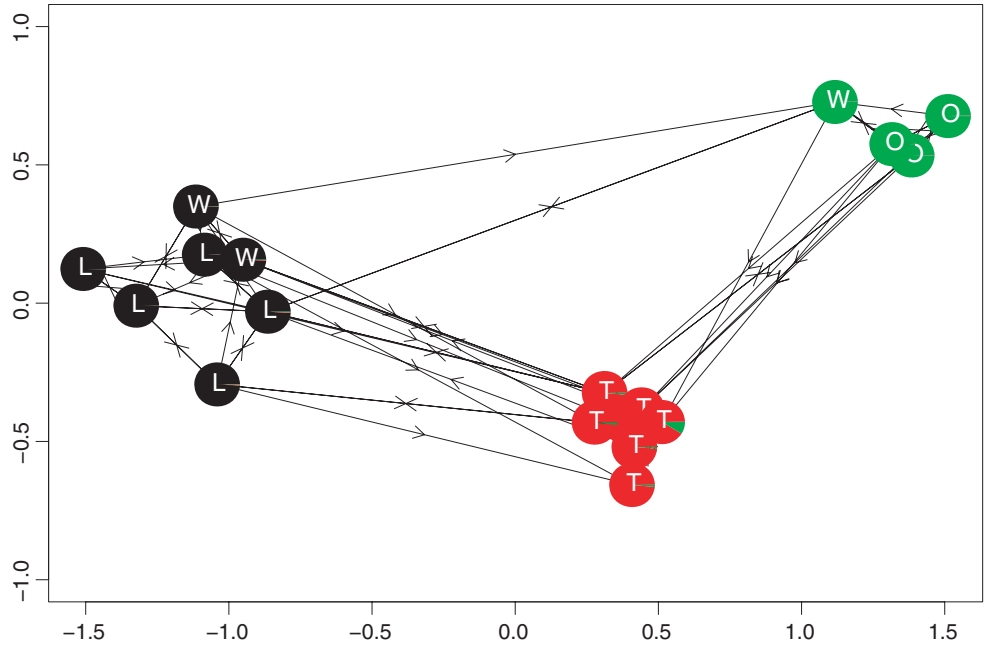

Fig. 3. Estimates of clusters and latent positions for the relationship between monks within a monastery from the Bayesian estimation of the latent position cluster model: the probability of assignment to each latent cluster is shown by a pie chart

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.

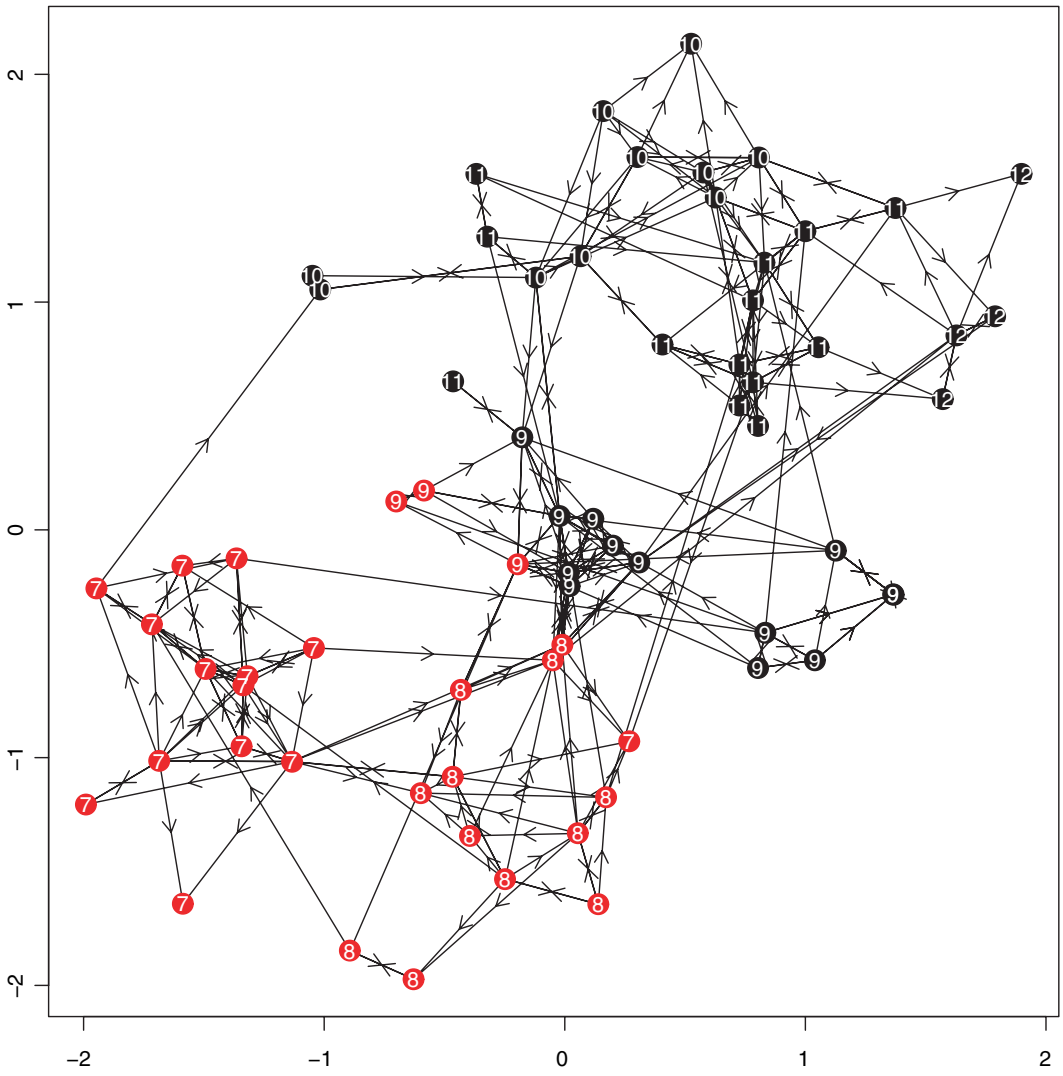

Fig. 4. Two-stage maximum likelihood estimates of the latent positions and clusters for the adolescent health data, where the number of clusters (2) is chosen by BIC: clusters are shown by colour with actual grades shown as numbers

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

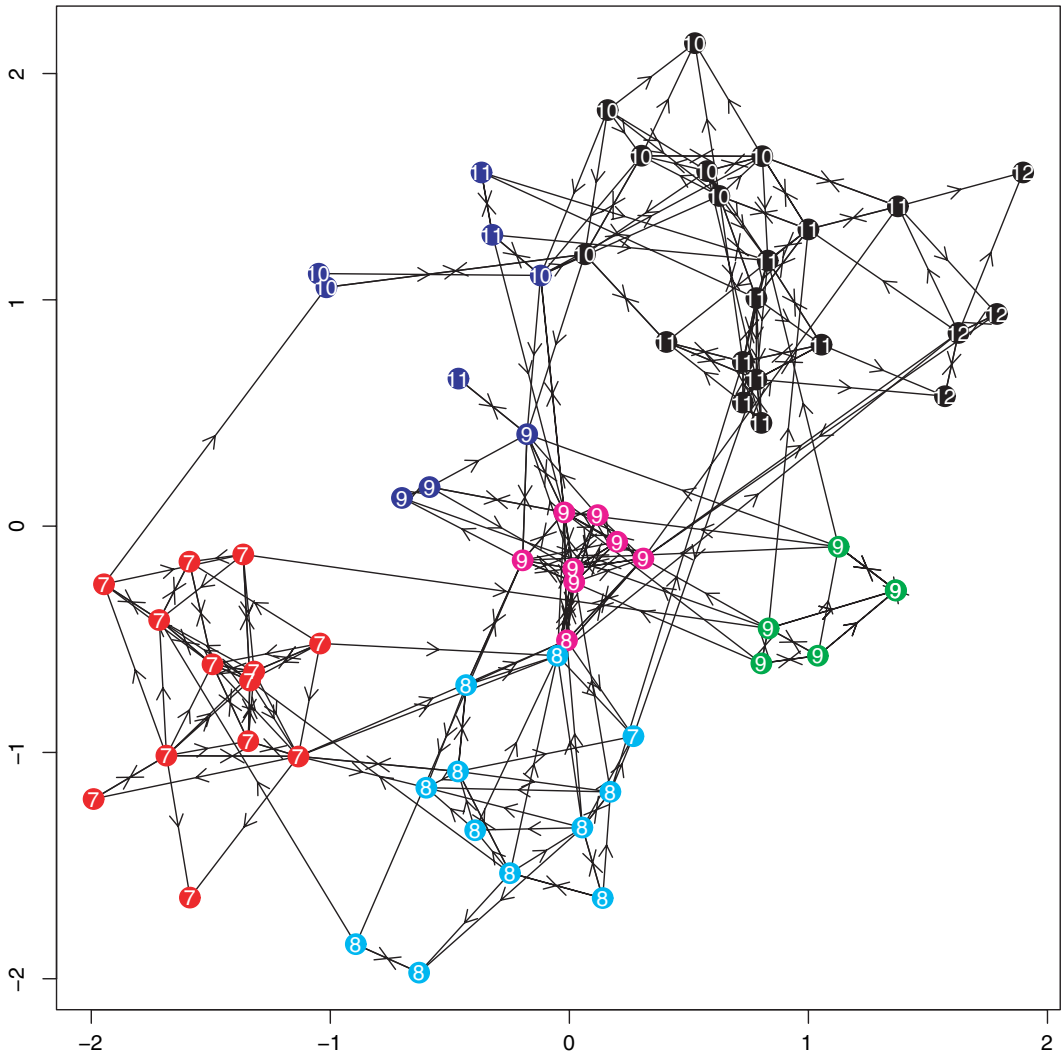

Fig. 5. Clusters from two-stage maximum likelihood estimates of the latent position cluster model for the adolescent health data, where the number of clusters is constrained to be 6: clusters are shown by colour with actual grades shown as numbers; there are six green points, representing students from 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 grades 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 50000 burn-in iterations that were discarded, and a further 2 million iterations, of which we kept every 1000th value. The result-

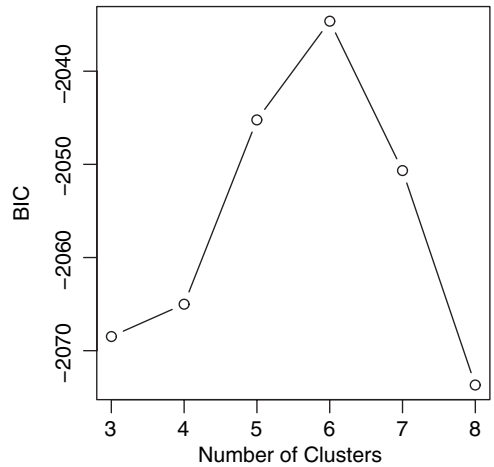

Fig. 6. BIC-plot for the latent position cluster model of the adolescent health network

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 to 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.

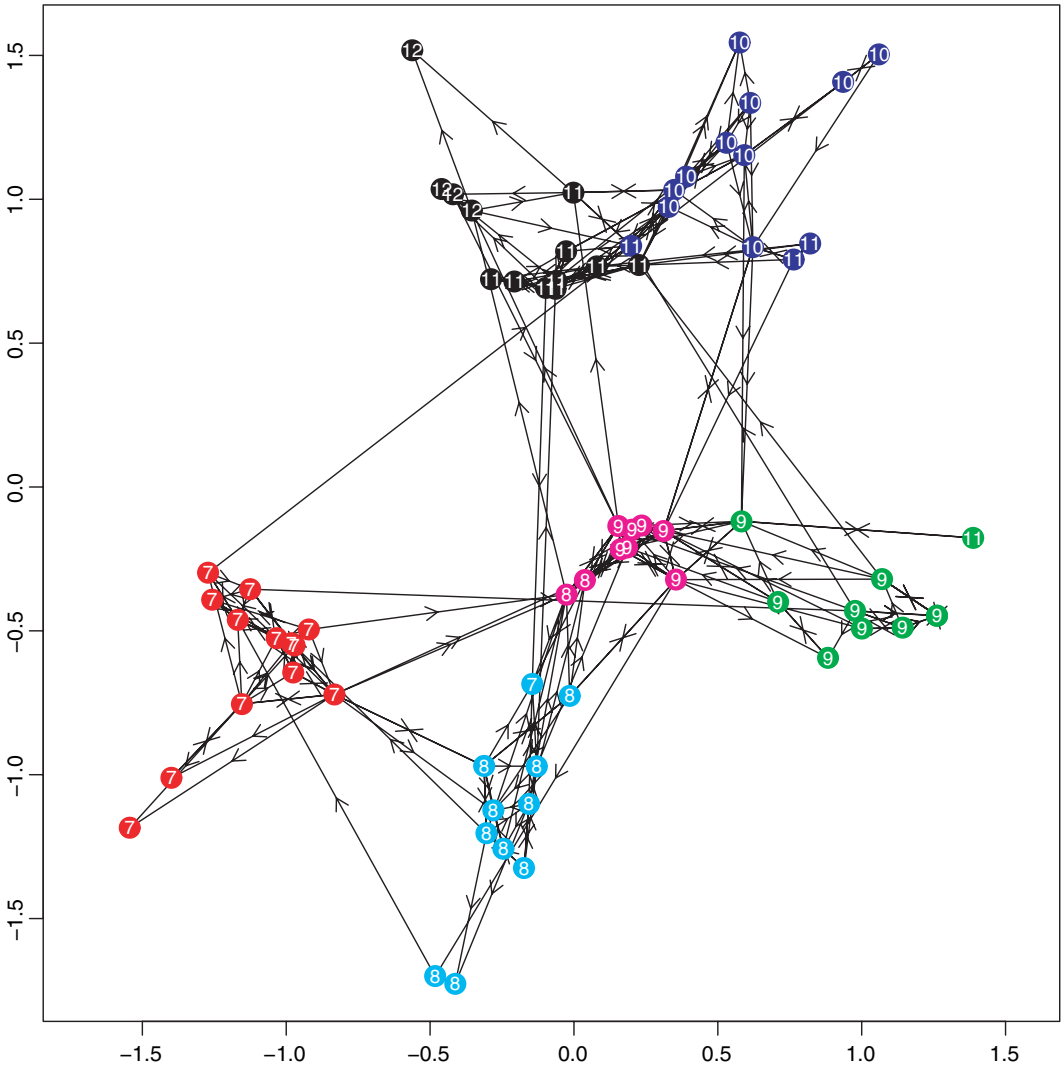

Fig. 7. Bayesian estimates of posterior clusters and latent positions for the friendship network in the adolescent health school: latent clusters are shown by colour with actual grades shown as numbers

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

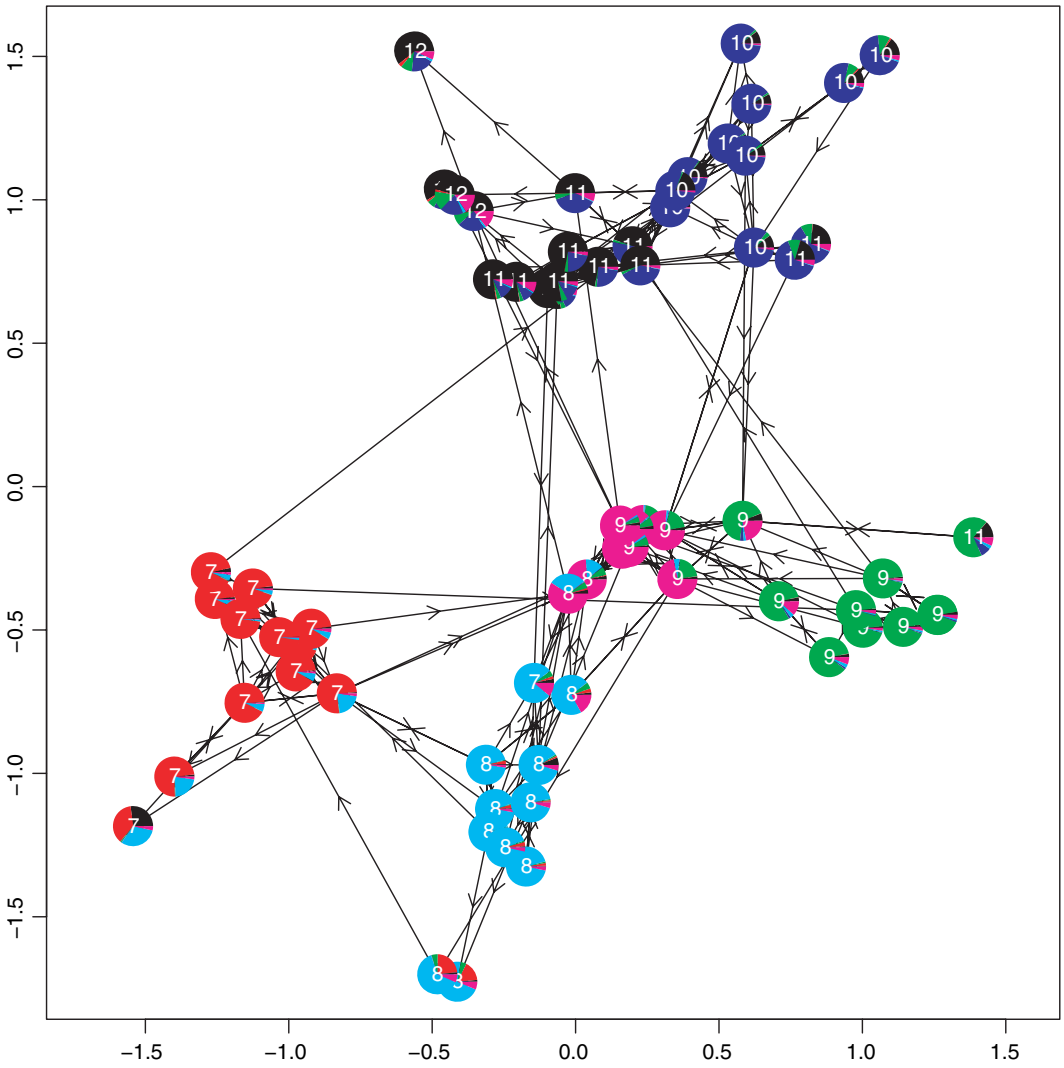

Fig. 8. Pie charts for posterior probabilities of cluster assignment for each actor, at the Bayesian estimates of posterior latent positions for the friendship network in the adolescent health school: the students' grades are shown as numbers

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` has been written to implement the procedures in this paper (Handcock *et al.*, 2004). The package 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 National Institute on Drug Abuse grant DA012831 and National Institute of Child Health and Human Development 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

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. Classifcn*, **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.
Handcock, M. S., Tantrum, J., Shortreed, S. and Hoff, P. (2004) latentnet: an R package for latent position and cluster modeling of statistical networks. (Available from `http://www.csde.washington.edu/statnet/latentnet`.)
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. E., Robins, G. L. and Handcock, M. S. (2006) New specifications for exponential random graph models. *Sociol. Methodol.*, **36**, 99–153.
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 Anderson, C. J. (1987) Stochastic a posteriori blockmodels: construction and assessment. *Socl Netwrks*, **9**, 1–36.
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.

Discussion on the paper by Handcock, Raftery and Tantrum

Tom A. B. Snijders (*University of Oxford and University of Groningen*)

The paper by Handcock, Raftery and Tantrum is an interesting new step in modelling social networks—more specifically, digraphs (directed graphs). This is a basic data structure for representing relational data, which is found to be increasingly important in all the social sciences.

To make a plausible stochastic model for one observation of a digraph, the crucial issue is to represent the stochastic dependence between the tie variables $Y_{i,j}$. Important types of dependence are dyadic dependence between reciprocal tie variables $Y_{i,j}$ and $Y_{j,i}$, and triadic dependence between tie variables involving three nodes, such as the pair $(Y_{i,j}, Y_{i,k})$, or the triple $(Y_{i,j}, Y_{j,k}, Y_{i,k})$, which is used to represent tendencies towards transitivity.

Two general ways for representing dependence between tie variables have been presented in the literature. One is by postulating latent nodal variables and conditional independence of the observations, given the latent variables, in the classical Lazarsfeld tradition of latent structure models. This way is followed in the present paper. A discrete latent class approach was proposed in Nowicki and Snijders (2001). The second way is by directly modelling this dependence, as is done in exponential random-graph models. The landmark paper here is Frank and Strauss (1986); this type of modelling has become practically feasible especially since the model extensions that were recently presented in Snijders *et al.* (2006). Further research is needed to compare these ways of representing dependence in stochastic digraphs theoretically and empirically; one very practical advantage of the latent structure models is that they allow us to handle randomly missing data in almost trivial ways—which is quite unusual for techniques of social network analysis.

When using a latent distance model, a major question is the type of metric to employ. The current paper proposes a Euclidean metric with a superimposed clustering. An alternative is an ultrametric (Schweinberger and Snijders, 2003) which is equivalent to a system of nested groups, or clusters, without further structure. When employing the Euclidean metric the modeller must choose the number of dimensions; for the ultrametric model the number of nesting levels. The Euclidean metric is richer and its spatial arrangement has effects on the probability of ties both within and between clusters. Filling in this richer detail also requires more from the data. Which metric is more appropriate is an empirical matter and also depends on the substantive background knowledge and the substantive questions being asked. From the empirical side, a more detailed assessment of fit would be interesting than only the Bayes information criterion approximations that are proposed in the paper as a global measure of fit. It seems a good idea to follow the

paper and to assess fit conditional on estimated positions (although this is not appropriate for comparing the fit with other types of model). Detailed fit assessments could be based on the contributions of dyads to the log-likelihood,

$$l_{i,j}(Z) = \log\{P(Y_{i,j} = y_{i,j}|Z, X, \beta)\} + \log\{P(Y_{j,i} = y_{j,i}|Z, X, \beta)\}$$

(ignoring in the notation for l the dependence on X and β). The fit of node i can be assessed on the basis of the sum $\Sigma_{j=1}^{n} l_{i,j}(Z)$. Comparing this across nodes will indicate which nodes conform relatively poorly to the Euclidean model. Similarly when C_g for $g = 1, \ldots, G$ are the sets of nodes defining the G clusters after post-processing the data, the quality of the representation of the within-cluster and between-cluster ties can be measured by using

$$\sum_{i,j \in C_g; i<j} l_{i,j}(Z)$$

and

$$\sum_{i \in C_g,\ j \in C_h} l_{i,j}(Z)$$

respectively (where $g \neq h$).

The sampling properties of such fit statistics will be quite complicated and a parametric bootstrap may be too time consuming to approximate them. However, given that we are discussing a fit assessment conditionally on the estimated latent positions, a natural first-order standardization is to treat the $Y_{i,j}$ as independent binary variables for the given X, Z and β, and to standardize by using the accordingly calculated means and variances.

Next to a detailed fit analysis, a detailed sensitivity analysis will be interesting, to assess how well the data determine the Euclidean positions. Denote by $Z[i ++ u]$ the array Z in which the position of node i has been translated by adding to it the vector u. Then $\Sigma_j l_{i,j}(Z[i ++ u])$ indicates the sensitivity of the conditional log-likelihood to translation of node i by the vector u. This will have a local maximum in $u = 0$ and preferably is approximately concave as a function of u. If there are regions in space that are away from $u = 0$ with local maxima which are not much lower than the value in $u = 0$, then node i has an ambiguous position. Similarly, the change in log-likelihood can be calculated that results from translating all nodes in a whole cluster C_g by the same vector u, or by orthogonally rotating all points in a cluster. This will yield possibilities for diagnosing how well the between-cluster patterns of ties determine the relative positions of the clusters.

The clustered spatial representation that is proposed in this paper represents what sociologists call the cohesive structure of the network. However, digraphs can have many structural features, and the cohesive structure is only one. While remaining in the framework of continuous latent variable models, it is straightforward also to represent the structural properties of hierarchy and of prominence. Hierarchy means that there is an order between the actors, and ties have a preferential direction. This can be important, e.g. when the relationship that is under study is an advice relationship, where the hierarchy could reflect expertise. Status differences also can give rise to hierarchically structured networks. Denoting now by z_{1i} instead of z_i the (multidimensional) locations representing propinquity, hierarchy can be represented by using an additional vector of one-dimensional latent variables z_{2i}, and adding

$$\beta_2(z_{2i} - z_{2j})$$

to the log-odds of the tie from i to j. This can be complemented by a third vector of latent variables, again one dimensional, contributing

$$\beta_3(z_{3i} + z_{3j})$$

to the log-odds of a tie; this represents prominence, defined as the propensity to have ties. For (Z_{2i}, Z_{3i}), we could postulate mixtures of (correlated!) bivariate normal distributions. This is nothing other than a reparameterization of the random activity and popularity effects of van Duijn *et al.* (2004). The parameterization that is proposed here has the advantage that it directly expresses the hierarchical aspect of the network structure, which is substantively interesting in many applications. When employing latent Euclidean distance models to represent directed social networks, it seems to me that the default should also be to include hierarchy and prominence (or activity and popularity) dimensions in the latent variables.

Also after such an extension, the latent space models and the exponential random-graph models currently are the main 'competitors' for statistically modelling non-longitudinal observations of social networks. Further practical experience with these models is necessary to assess their worth; this will need to involve more detailed studies of fit and sensitivity than we have seen so far. I expect that, especially for modelling larger networks (with, say, a few hundred or more nodes), the latent space models will not be able to represent network structures as expressed by subgraph counts sufficiently well and the exponential random-graph models will not be able to represent the cohesive structure sufficiently well. Models that combine important features of these two approaches may be the next generation of social network models.

I am very pleased to propose the vote of thanks for this very interesting paper.

Tony Robinson (*University of Bath*)

A compelling advantage when using mixture model-based clustering in a Bayesian framework is the ability to obtain posterior probability information on *all* quantities of interest. This paper neatly incorporates the technology in a novel approach for social network discovery. However, the benefits of mixture modelling come with a cost, especially in a clustering application for, by definition, objects are 'mixed up' and the results of analyses must be examined carefully to glean the important information about likely forms of object partitioning. This is certainly so when clustering on observables and now, challengingly, we see the authors applying the technique to positions of actors in a social space which is latent. Partitioning these actors is clearly a major objective in the current exercise. Most clustering methods partition reasonably well when the degree of separation between groups is marked and correspondingly less well as the degree of overlap increases and model-based clustering is no exception. In using model-based clustering we often find that the results for marginal quantities often conflict or obscure. For example there may be a tension between the likely number of components and the sampled frequency of likely partitions. A careful examination of the marginal, joint and conditional behaviour of the results is necessary for sensible inferences to be drawn.

Moreover the results of model-based clustering can be sensitive to structural assumptions which directly affect sampled partitioning. One such structural assumption here is that of spherical Gaussian components. The authors give some justification for such a choice based on invariance of the likelihood under the co-ordinate system and sphericity will certainly conveniently cut down on the number of parameters. But I do not find this wholly convincing and wonder whether the authors are truly averse to alternatives such as Gaussian components with a more flexible covariance structure or even non-Gaussian components. Are there underlying substantive considerations concerned with the nature of the social space and the clustering behaviour of actors within it or is social space so flexible as to render the distributional model choice essentially immaterial? I doubt that the latter is always so and question whether such a restrictive model can sometimes lead to too many clusters or overdispersed estimates which would affect the partitioning.

The authors take a fairly traditional approach to deciding on the number of clusters apart from conditioning on actor positions. There are other approaches such as transdimensional samplers but they require even more sophistication in implementation. If determination of the number of clusters is made conditional on a posterior estimate of actor positions, care must be taken to ensure that these are determined fairly and are not influenced by an overly restrictive model specification and by design of a sampler that allows a free mixing of actor positions across the latent space to avoid imposing artificial clustering. Deciding on the number of clusters needs to take account of all aspects of the model.

This determination of configurations of actors in the social space has clear parallels with multidimensional scaling in that both techniques aim to produce an interpretable configuration in a space of specified dimension from which structure can be identified. The default choice for the dimension of the latent space seems to be 2 as it is in most applications of multidimensional scaling undoubtedly driven by ease of visualization in both cases. It would be standard good practice in multidimensional scaling to explore solutions in other nearby dimensions and I would have liked to see the same in the two examples of latent space clustering. The authors do reference Oh and Raftery (2001) as a possible way to choose but otherwise leave the choice of dimension to be specified by the user who will no doubt also choose 2 as a starting- and possibly a finishing point. For example in the adolescent health example, would the choice of three or more dimensions lead to separation of the higher grades? Similarly would a choice of one dimension yield essentially the same results as two dimensions, as inspection of Fig. 8 seems to indicate groups with roughly increasing grades with anticlockwise movement around the configuration and the higher grades curling back towards the lower.

I also have a worry about the basic underlying model as specified in equation (2). If we accept that there may be underlying clusters, why should the covariates not act differently among them? The global behaviour in equation (2) clearly does not allow this possibility.

I believe that the authors have made a decent start at designing a potentially useful technique for clustering in static social networks but that users need to be aware that the technique is far from problem free and that they must be careful not to overcook the recipe and thereby to overinterpret results. It is my pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

Anthony C. Atkinson (*London School of Economics and Political Science*) **and Marco Riani** (*Università di Parma*)

Over the years we have enjoyed both Adrian Raftery's talks and his flow of publications on model-based clustering. We would like to compare some results of the use of `mclust` with a cluster analysis that is produced by the use of the forward search.

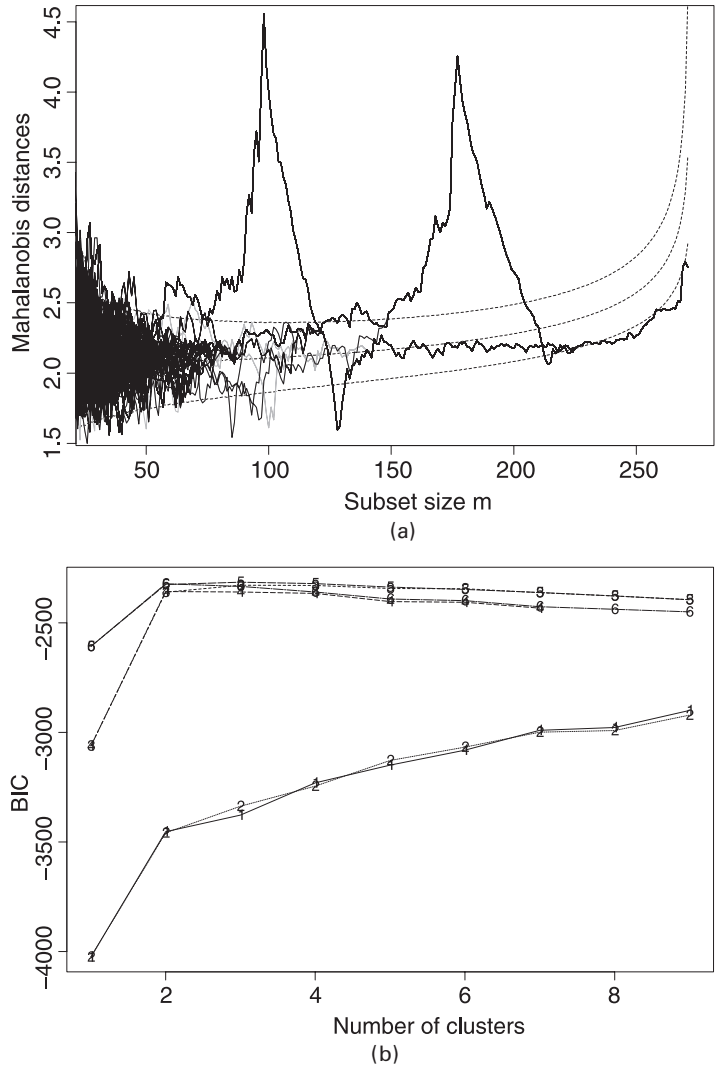

Fig. 9. Old Faithful Geyser data: (a) forward plot of minimum Mahalanobis distances from 300 random starts with 1%, 50% and 99% envelopes (two clusters are evident); (b) Bayes information criterion plot from mclust, indicating three clusters

The forward search for multivariate data is described in Atkinson *et al.* (2004). In general the search proceeds by successively fitting subsets of the data of increasing size. For a single multivariate population any outliers will enter at the end of the search with large Mahalanobis distances. If the data are clustered and the search starts in one of the, unfortunately unknown, clusters, the end of the cluster is indicated when the next observation to be added is remote from that cluster. To find clusters we have recently (Atkinson *et al.*, 2006a, b; Cerioli *et al.*, 2006) suggested running many searches from randomly selected starting-points. Some of these start in, or are attracted to, a single cluster; a forward plot of the minimum Mahalanobis distance of the observations that are not in the fitted subset then reveals the cluster structure.

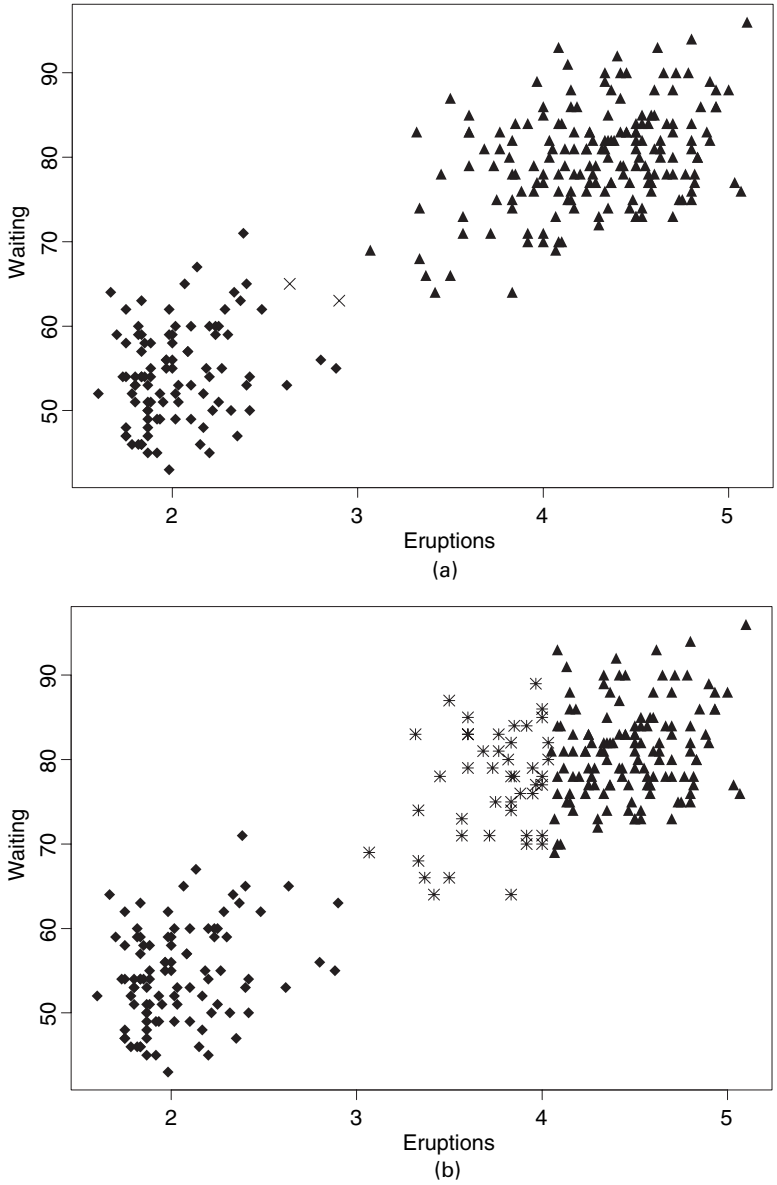

Fig. 10. Old Faithful Geyser data: scatterplot matrices of (a) the two clusters from the forward search (units marked × could lie in either cluster) and (b) the three clusters that were found by mclust

As an example we analyse 272 observations on the eruptions of the Old Faithful Geyser taken from the Modern Applied Statistics in S library (Venables and Ripley, 2002). Azzalini and Bowman (1990) described the scientific problem. Fig. 9(a), a forward plot of Mahalanobis distances from the forward search, clearly shows the two groups. Fig. 9(b) is the Bayes information criterion output of `mclust` from S-PLUS which, on the contrary, indicates three clusters. Fig. 2 of Fraley and Raftery (2006) for a slightly different set of geyser data is similar and again indicates three clusters.

We use further forward searches to establish membership of these two clusters and establish the unclustered units. A scatterplot of the resulting two clusters is shown in Fig. 10(a). The three clusters that are found by `mclust` are in Fig. 10(b). Fuller details of our analysis including further comparisons and considerations of robustness are in Atkinson and Riani (2007).

We do not want to imply that the imposing edifice, some of whose rooms we have so enjoyably visited today, is built on sand. But it does seem that there are still some fundamental problems in the foundations of clustering that need to be resolved.

Isobel Claire Gormley (*University College Dublin*) **and Thomas Brendan Murphy** (*Trinity College Dublin*)
We congratulate the authors on a thought-provoking paper. We feel that the combination of model-based clustering with latent space modelling is applicable far beyond the proposed application of the analysis of social network data.

We have recently been developing statistical models for rank data including data from Irish elections (Gormley and Murphy, 2005, 2006a) and Irish college applications (Gormley and Murphy, 2006b). Two approaches that we have taken are using mixture models (Murphy and Martin, 2003; Gormley and Murphy, 2005, 2006b) and using a latent space model (Gormley and Murphy, 2006a). The paper that is presented here provides a combination of both of these modelling approaches which we look forward to applying to the analysis of rank data.

In our work, we found that model choice is a difficult aspect of the modelling process. The number of components in a mixture model can be estimated consistently by using the Bayes information criterion (Keribin, 2000) but we found that the choice of dimensionality in our latent space model is more problematic. We were wondering whether the authors could provide us with insight into the methods for choosing the dimensionality of the latent space in their social network model.

More recently, we have considered methods for including covariates in our models. One approach that we have considered is allowing the mixture probabilities to depend on covariates (Gormley, 2006); this yields a special case of the mixture-of-experts model (Jacobs *et al.*, 1991). This model can be fitted very easily with minor changes to the mixture modelling framework. In the context of this paper this may provide an alternative method for achieving homophily by attributes.

Trevor Sweeting (*University College London*)
I would be interested to hear from the authors whether they have considered using an infinite group cluster model and, if so, what they would consider to be the relative advantages and disadvantages of such a formulation over their finite group cluster model in the context of network models. There are various possible Bayesian formulations of infinite group cluster models. A common choice of prior distribution for the group weights λ arises from a Dirichlet process prior structure for the parameters of the latent positions, since this structure automatically induces clustering. Specifically, writing $\phi_i = (m_i, s_i^2)$ the Dirichlet process mixture (DPM) structure for the latent positions would be specified as

$$z_i|\phi_i \sim \mathrm{MVN}_d(m_i, s_i^2 I_d),$$

$$\phi_i|F \overset{\mathrm{IID}}{\sim} F,$$

$$F \sim \mathrm{DP}(F_0, \gamma),$$

$$F_0 = \mathrm{MVN}_d(0, \omega^2 I_d) \times \sigma_0^2 \, \mathrm{Inv}\chi_\alpha^2.$$

Here $\mathrm{DP}(\cdot, \cdot)$ denotes a $(d+1)$-dimensional Dirichlet process and F_0 and γ are the associated mean and precision parameters. Now, for $g = 1, 2, \ldots$, write $\theta_g = (\mu_g, \sigma_g^2)$ and $\theta = (\theta_1, \theta_2, \ldots)$. Using Sethuraman's (1994) stick breaking representation of the Dirichlet process, the above specification is equivalent to the following infinite group version of the authors' model:

$$z_i|\theta \overset{\mathrm{IID}}{\sim} \sum_{g=1}^{\infty} \lambda_g \, \mathrm{MVN}_d(\mu_g, \sigma_g^2 I_d),$$

$$\theta_g \overset{\text{IID}}{\sim} F_0,$$

$$\lambda_g = u_g \prod_{j=1}^{g-1} (1 - u_j),$$

$$u_j \overset{\text{IID}}{\sim} \text{beta}(1, \gamma).$$

For additional flexibility a prior distribution is often assigned to the precision parameter γ, chosen to reflect the prior expectation and uncertainty about the number of clusters that are contained in the data. It would be of interest to explore whether Markov chain Monte Carlo (MCMC) schemes in the literature in the case where the z_i are not latent (see, for example, Neal (2000)) could be readily integrated with the MCMC scheme that is given in Section 3.2 of the paper.

The DPM model is just one possible model for Bayesian clustering. The generalization of the DPM model to product partition models, for example, is described in Quintana and Iglesias (2003). Potential advantages of an infinite group over a finite group cluster model would be that, firstly, neither the Bayes information criterion nor reversible jump MCMC methods would be necessary for estimation of the number of clusters, G, contained in the data; secondly, uncertainty about G could be readily assessed; thirdly, an infinite group model would deal more cleanly with the situation that is discussed in Section 6 where network data are available for only part of a population so that other clusters may not yet have been represented.

David S. Leslie (*University of Bristol*)

I congratulate the authors for their interesting paper. However, it seems that the Markov chain Monte Carlo sampling scheme that was used results in extremely slow mixing, requiring 2 million iterations with only every 1000th iteration being used. One aspect of this slow mixing relates to a problem that was encountered by Leslie *et al.* (2006).

The problem arises when we move from a simple latent structure to a mixture model latent structure. In Leslie *et al.* (2006) the transition was from simple probit regression with a normal latent structure to a binary choice regression model with latent variables drawn from a Dirichlet process mixture model. In the current paper the transition is from the simple normal model for the latent variables that was used by Hoff *et al.* (2002) to a situation in which the latent variables are drawn from a mixture of multivariate normal distributions. The natural sampling scheme to use when such an extension is made is that presented in the paper, where the component labels K are sampled conditionally on the latent variables Z; then the latent variables are sampled conditionally on the component labels K. However, a simple example suffices to see that this is likely to result in extremely poor mixing.

Consider the two-dimensional latent variables that are shown in Fig. 11, and consider first the process of updating the latent variable of the point that is marked; we shall call it point i. The latent variable z_i is drawn conditionally on membership of cluster 1 and so is highly likely to be close to the other members of cluster 1, and hence far from the members of cluster 2. Now consider updating the cluster label K_i conditionally on the latent variable value z_i: it is highly unlikely that i will be allocated to cluster 2 owing to the location of the latent variable z_i. As seen by this example, it is very difficult for the latent variables to move between clusters, owing to the high correlation between cluster labels K_i and latent values z_i.

The solution that was proposed by Leslie *et al.* (2006) is, for each i, to update K_i and z_i simultaneously. They could integrate out a single z_i, allowing Gibbs sampling of cluster membership without conditioning on the latent variable, followed by Gibbs sampling of the latent variable conditionally on the sampled cluster membership. In the situation that is presented here it seems to be impossible to perform this trick,

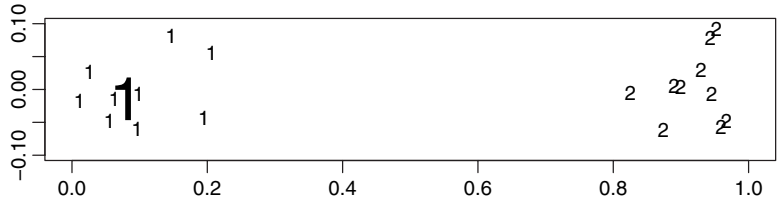

Fig. 11. Latent variables z, with cluster labels indicated by numbers and point i indicated by the large number 1

but nevertheless we can propose K_i^* at random, and then propose a value of z_i^* conditional on K_i^*, before deciding whether or not to accept both K_i^* and z_i^* by using a Metropolis–Hastings ratio.

N. T. Longford (*SNTL, Leicester, and Universitat Pompeu Fabra, Barcelona*)
Mixtures of multivariate normal distributions, which are used by the authors with great skill, are a greatly underrated device for generating a wide variety of distributions. In more than two dimensions, the normal is the only comprehensive class of distributions that is easy to handle in the standard likelihood-related calculations. By fitting a multivariate normal mixture we approximate the target distribution. We are fitting not only the modes of the distribution but also its shoulders and tails. Therefore mixture components cannot be automatically associated with clusters. All clusters are 'condemned to be normal'. For example, none of the clusters in the authors' examples could be associated with a skewed latent distribution. Admittedly, the layer of latentness grants some flexibility in this respect, but a mixture component can be declared a cluster only when its variances are small relative to the distance of its expectation from the expectations of the other components. In a different context, Longford and Pittau (2006) present an analysis in which multivariate mixtures cannot be regarded as clusters because the mixture components differ principally by their patterns of variation and dependence.

The term 'determination' (of the number of components) sits very uncomfortably in the Bayesian terminology, because it implies elimination of any uncertainty. Instead of concluding that there are three components in the first example, I would prefer a brief discussion of the solutions for two and four components, accompanied by a comment on how much less likely they are and how their solutions are related to the preferred three-component solution.

I understand that inferences in both examples are made for the subjects in the respective studies, not for the underlying population. In a frequentist view, there is a puzzling ambiguity. What distribution would govern the responses of the 18 monks in a replication of the study: the posited model? A more realistic alternative is that some (strong) links would be declared in every replication, whereas some other (weaker) links may be declared with a range of probabilities. If all the links are strong there is no variation in the replicate response patterns and, presumably, there is no uncertainty in the inference. Would the process of forming links be also replicated? I would be concerned if these issues were regarded in the Bayesian paradigm as not relevant, even though I appreciate that some would be difficult to incorporate in the analysis.

John T. Kent (*University of Leeds*)
The motivation for the statistical models in this paper is mainly focused on the networking and social sciences perspective. However, it is also helpful to draw out the analogies to more conventional statistical methodology. For example, in regression analysis, we can

(a) start with a geometric relationship $y = a + bx$,
(b) include normal errors to obtain the usual linear model,
(c) extend this framework to a generalized linear model, with for example Bernoulli observations,
(d) include random effects to allow for grouping and
(e) view clustering as an unlabelled random-effects model.

Similarly, for data on the relationships between n individuals or sites, we can

(a) start with the mathematical result that knowing all the Euclidean distances between the n sites determines the configuration of sites (up to translation and orientation),
(b) introduce stochastic errors to obtain the classic multidimensional scaling estimation problem,
(c) extend this framework to a generalized linear model for the presence or absence of edges and
(d) introduce random effects and
(e) introduce clustering as before.

Since an underlying latent configuration is determined only up to translation and orientation (and often size), it enters the statistical model only through its shape. For example in equation (7), when updating μ_g, it is only the shape (or shape plus size) of the configuration which is identifiable, not the whole configuration, and the updating exercise should take this restriction into account. I suspect that the incorporation of shape ideas into the analysis would have only a minor practical effect, but it would be the 'right' approach in terms of identifiability.

A recent development in shape analysis is the investigation of unlabelled shapes, where we may want to match two configurations together, but not know which sites correspond. One application is to protein

structure analysis, where the positions of atoms on each protein can be determined by X-ray crystallography and where it is suspected that (subregions of) proteins of similar shape have similar biological function, but where the labellings are unknown. There are some similarities to the problem of comparing clusterings determined by the $\{\mu_g\}$ in different Markov chain Monte Carlo simulations.

Tony Lawrance (*University of Warwick, Coventry*)
It is a pleasure to contribute to the discussion after the experts have spoken. My experience of this area and paper is the 2-hour train journey from Warwick to London, a distance of nearly zero according to the metric which this paper induced. My first point is to enquire how the analysis can address the measurement of friendliness or interactions of the actors that are involved, rather than their groupings. Secondly, I noticed that the modelling is predicated on a conditional independence assumption, which it must be tough to validate and is probably a matter of faith. I could not immediately see any attention in the paper given to assessing the fit of the model, and the choice of prior forms seemed clever, but I wonder how much can they influence the final groupings? Wider empirical validation, replacing monks and monasteries by lecturers and departments, would satisfy me more generally. My final observation concerns the microscopic pie charts, noting that they take the development of invisible graphics to a new level, at least judging by my monochrome preprint. Overall, I thought that this paper was a nice blend of methodology and application.

The following contributions were received in writing after the meeting.

Edoardo M. Airoldi (*Carnegie Mellon University, Pittsburgh*)
The authors' work with the *latent space clustering* methodology provides an impressive demonstration of the use of hierarchical models for identifying groups of nodes from observed connectivity patterns. Modelling choices based on sociological principles, i.e. transitivity and homophily, increase its appeal as an exploratory tool for the analysis of social networks. The methodology proposed goes only part way, however, towards addressing fundamental issues that arise in the statistical analysis of social networks.

The *stochastic blockmodel of mixed membership* in Airoldi (2006) and Airoldi *et al.* (2007a) offers an alternative approach with different insights on latent aspects underlying network structure. Models in this family also posit the existence of an unknown number of clusters; however, they replace latent positions with mixed memberships $\pi_{1:N}$, which map nodes to (one or more) clusters, and add a *latent blockmodel* B that specifies cluster-to-cluster hierarchical relations. These parameters are directly interpretable in terms of notions and concepts that are relevant to social scientists, and better suited to assist them in extracting substantive knowledge from noisy data, ultimately to inform or support the development of new hypotheses and theories. Therefore, inference about $\pi_{1:N}$ and B is crucial for the analysis of data.

Applying this to Sampson's data demonstrates both linkages and differences. Our version of the Bayes information criterion also suggests the existence of three factions among the 18 monks, but our groupings are different. In Fig. 12, Romul and Victor (two of Sampson's Waverers) stand out; and so do Greg and John who were expelled first from the monastery. The mixed membership map is specified by using node-specific latent vectors $\pi_{1:18}$, independent and identically distributed samples from a three-dimensional symmetric Dirichlet(α) distribution. The map of hierarchical relationships among factions is specified by a 3×3 matrix of Bernoulli hyperparameters B, where $B(i, j)$ is the probability that monks in the ith faction relate to those in the jth faction. Other features that are relevant to data analysis are the marginal probability of a relation ($\pi_n' B \pi_m$) and the relation between the number of clusters and dimensionality of the latent simplex.

Our models allow a focus on issues such as membership of monks in factions, and this could lead to the formation of a social theory of failure in isolated communities, which is capable of testing with longitudinal data. In Airoldi *et al.* (2007), we provide full details on specification, estimation and interpretation for both the Sampson and the adolescent friendship network examples.

Julian Besag (*University of Washington, Seattle*)
I would like to comment on the authors' choice of examples. After all, social networks have been around for a long time and there is an abundance of data, so we should be expecting more than purely illustrative analyses by now.

In their first example, the authors deem an edge from i to j to exist if i cites j at *any* of his three interviews. In general, if clusters change over time, such a rule could lead to spurious results. Moreover, as regards social science, I would assume that the temporal development of clusters, including their creation, coalescence, fragmentation and destruction, is of more interest than their static properties. Although three

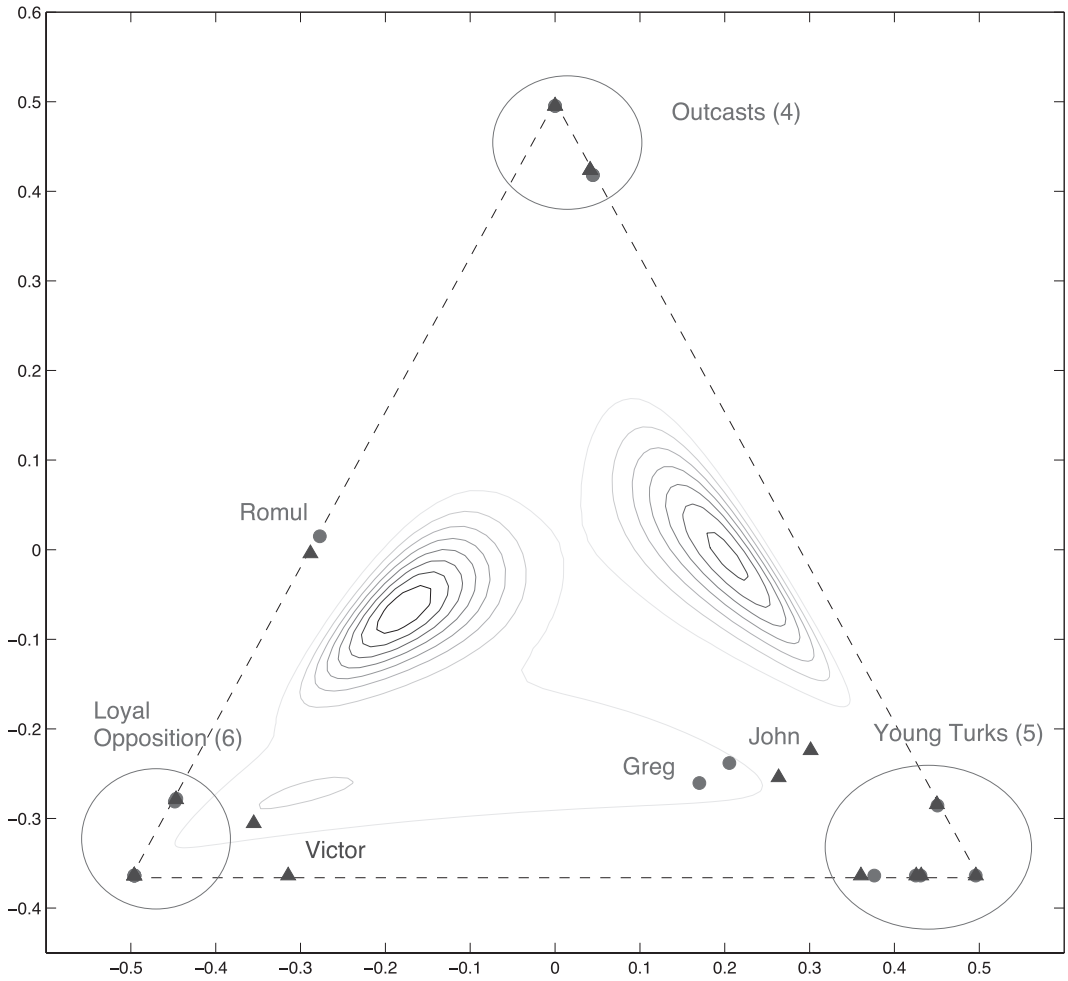

Fig. 12. In the reference simplex, circles and triangles correspond to mixed membership vectors of individual monks, $\pi_{1:18}$ (circles were obtained with $B = \mathrm{Id}_3$ and $\alpha = 0.01$, whereas triangles were obtained with $B = \mathrm{Id}_3$ and $\alpha = 0.58$—estimated via an empirical Bayes method): an arbitrary one-to-one projection situates the Gaussian mixture of Table 1 in the simplex

time points are probably too few for meaningful analysis, more extensive space–time networks could have been chosen. Such analysis is particularly important for communicable diseases. Note that, in setting up space–time models, multiple changes in edge configurations can occur (almost) instantaneously, though this is sometimes overlooked.

As regards their second example, do the authors have a justification for focusing on one particular school out of 132? It seems to me that they should at least have analysed a small sample of schools. And why were no covariates included, particularly the grade of student? To claim success in extracting grade as an important clustering attribute suggests to me that the authors are too easily satisfied. Their secondary conclusions are plausible and could have been checked in other schools. The general point here is the effect of including cluster attributes as covariates, which is allowed in their original formulation but apparently not in their examples. How does this affect cluster identification?

Lastly, do the authors have anything to add about the relevance of their approach to the huge networks that for example AT&T and Microsoft researchers must deal with and for which quite different methods are used? Is this merely a computational issue or is it that exploratory techniques are more appropriate?

David Blei (*Princeton University*) **and Stephen E. Fienberg** (*Carnegie Mellon University, Pittsburgh*)

We congratulate Handcock, Raftery and Tantrum for this interesting and elegant paper that proposes combining the latent space and stochastic blockmodels of sociometric data. We found it especially instructive since it parallels our efforts to develop a similar analysis (Airoldi *et al.*, 2007a, b). We shall compare the two approaches.

The authors' construction integrates the latent space model for relational data with a 'traditional' cluster model based on a finite mixture of Gaussian distributions. Their methodology mixes the Bayesian approach to cluster estimation with a likelihood variant of the latent space model. This is valuable for exploratory analyses of sociomatrices.

Our approach begins with a random mixed membership vector for each actor (Erosheva, 2003, 2004; Blei *et al.*, 2003). These vectors can be viewed as describing a soft clustering, where each actor belongs to multiple clusters with different proportions. The binary relationships between actors, i.e. the observed data, are mediated by per-pair latent variables, each drawn conditioned on an actor's mixed membership vector. In its general form, we allow for multiple relationships and covariates. This is a Bayesian hierarchical model.

The model that is proposed here can also be thought of as a hierarchical model, specifically when a Gaussian prior is placed on the latent position variables. In contrast, however, each actor belongs to a single cluster and the corresponding partition governs the observed relationships. There can be variance in the latent position variables, but the idea of belonging to two or more groups cannot be represented. Posterior uncertainty about cluster membership (depicted by the pie charts in the authors' figures) is different from *mixed membership*, which carries with it an additional level of uncertainty. That said, the latent space of the authors is quite comparable with our proposed space of cluster proportions. They map actors to Euclidean space; we map actors to the simplex.

We and the authors have the same goal: infer the underlying latent structure from an observed sociomatrix. In the mixed membership model, full Markov chain Monte Carlo sampling for any but the simplest problems is unreasonably expensive. We have appealed to variational methods for a computationally efficient approximation to the posterior. These methods can scale to large matrices because of the simplified approximation (but at an unknown cost to accuracy). It would be interesting to understand computational trade-offs for the authors' method as the sample size grows and when large numbers of covariates are added.

Ronald Breiger (*University of Arizona, Tucson*)

As Handcock and his colleagues refer to their model (in Section 1) as 'a stochastic blockmodel', and as they apply their latent position cluster model (LPCM) to a data set that was analysed much earlier by White *et al.* (1976) in their paper on blockmodels, it may be instructive to focus on the agenda that was put forward in the earlier paper and on the extent to which the new paper furthers that agenda.

As indicated in their paper's title, White *et al.* (1976) insisted on modelling social structure from 'multiple networks'. In blockmodel analysis, partitioning of individuals is only one side of a dual problem, the other being interpretation of the pattern that is formed by that partition. The clique pattern of sociometry, which seems well generalized by transitivity, homophily on attributes and clustering, as in the LPCM model, is only one possible pattern for a blockmodel, in which some sets of actors might be understood as structurally important because they have no ties among themselves but are all tied to the same other groups. A fundamental concern was modelling the catenation of ties of different networks (such as 'friends of advisers').

Recent lines of research have resulted in breakthroughs in carrying forward this agenda. Generalized blockmodelling (Doreian *et al.*, 2005) permits ideal block types defining network equivalence to differ across pairs of blocks. Exponential random-graph modelling is providing a firm statistical foundation for studying catenation of ties across multiple networks (e.g. Lazega and Pattison (1999), pages 84–85). And stochastic blockmodelling, which was developed for a partition of actors specified *a priori* (Wang and Wong, 1987) or on the basis of a clustering algorithm (the models that were reviewed in Section 1 of this paper), is supplying a firm statistical foundation to replace the *ad hoc* clustering procedures that were often used in the earlier work. The LPCM is so appealing because it is based on a specifiable model of network structure (though I hope that other models will eventually also be specified), because it articulates so well with the statistical foundations of exponential random-graph modelling, and because the results are so sharp. Social networks researchers are in debt to Handcock and his colleagues for these substantial advances.

Statistical work on blockmodels has focused on partitions of actors, but not yet on specifying patterns of equivalence among blocks. Future work might address this issue along with partitions across multiple

networks and those based on wider varieties of patterns of tie (such as those found in negative affect networks).

Carter T. Butts (*University of California, Irvine*)

Handcock, Raftery and Tantrum have ably demonstrated the potential for latent space models to address certain time-worn questions in network analysis within a modern statistical framework. One important limitation of this model, however, is that it cannot represent systematic biases in the orientation of asymmetric dyads (apart from covariate and/or activity effects). This is a simple consequence of the symmetry of $|z_i - z_j|$ and similarly holds for the projection model of Hoff *et al.* (2002). The inability to represent orientation bias is consequential in various settings, particularly where status differences are present. A natural example with relevance to the present paper would be systems of ranked clusters (Davis and Leinhardt, 1972; Holland and Leinhardt, 1970), in which we observe multiple cohesive social groups whose intergroup connections form a partial order. Biased asymmetry is also central to the incidence of transitivity *per se* (as opposed to mere triadic clustering), a feature which the authors single out as being of particular importance.

A simple extension which would rectify this limitation is suggested by the geographical literature on flow matrices. Given a matrix Y of point-to-point flows, Tobler (1976, 2005) suggested decomposing the matrix into symmetric ($Y^+ = (Y + Y^{\mathrm{T}})/2$) and skew symmetric ($Y^- = (Y - Y^{\mathrm{T}})/2$) components. The symmetric component matrix Y^+ is modelled via multidimensional scaling methods, whereas the skew symmetric matrix Y^- is modelled via a *potential surface* f, such that $\mathbf{E}(Y^-_{ij}) \propto f(v_i) - f(v_j)$. Intuitively, overall interaction is then governed by proximity in the latent space, whereas the direction of any asymmetries is determined by relative potential (with flow tending to proceed 'downhill' on the potential surface).

Incorporation of this notion into the authors' model is easily accomplished via modification of equation (2). Let $W = \{w_i\}$ be a set of latent vertex potentials, with $w_i \in \mathbb{R}^k \ \forall i$, and let β_2 be a k-vector of non-negative real parameters. We then posit that all edges are conditionally independent, with log-odds given by

$$\mathrm{logit}^{-1}(y_{ij} = 1 | z_i, z_j, w_i, w_j, x_{ij}, \beta) = \beta_0^{\mathrm{T}} x_{ij} - \beta_1 |z_i - z_j| + \beta_2^{\mathrm{T}}(w_i - w_j).$$

In many circumstances, it seems reasonable to assume $k = 1$ (i.e. a single-status ordering). $k > 1$ is possible, however, if multiple status dimensions are active within the network. The prior structure for W may be constructed analogously to that of Z, although the set of invariances is somewhat more restricted. The addition of vertex potentials to the latent space model is thus a very simple extension, but one which corrects a consequential limitation of the present approach.

Patrick Doreian (*University of Pittsburgh*) **and Vladimir Batagelj and Anuška Ferligoj** (*University of Ljubljana*)

The paper offers an intriguing approach to partitioning networks where the goal is to partition the vertices. Our comment points to an alternative approach: one that we hope is compatible with theirs.

Generalized blockmodelling (Doreian *et al.*, 2005) has a primary goal of discerning the structure of a network via homomorphisms of the network to simpler images. This entails explicitly partitioning both the vertices into clusters (called *positions*) and the relational ties into *blocks*. Blocks are specified by predicates that are used to characterize permitted block types that describe structure. Sets of predicates correspond to specific equivalences. For example, null and complete blocks correspond to structural equivalence whereas null and one-covered blocks correspond to regular equivalence.

The pattern of ideal blocks in the image characterizes the structure of the image matrix which, in turn, describes the underlying structure of the empirical network. Specifying a blockmodel can range from specifying only the permitted block types to specifying a block type for every location in a blockmodel. Given a specified blockmodel, empirical blockmodels are identified by a local optimization clustering algorithm that minimizes a criterion function: one that must be compatible with, and sensitive to, the equivalence that is defined for the specified blockmodel.

In the paper, transitivity is the driving structural feature. However, many blockmodels are consistent with transitivity. These include complete diagonal blocks with null blocks elsewhere, a complete upper triangular network with null blocks elsewhere (dominance structures) and complete diagonal blocks, for positions i and j with the upper (i, j) block complete and null blocks elsewhere. This suggests that transitivity is ambiguous with regard to block structures and incomplete for specifying network structure.

The likely structure of a high school network is one with denser patches of ties within grades. Six grades suggest six positions and a blockmodel structure of diag(den) where a density threshold is set. The off-

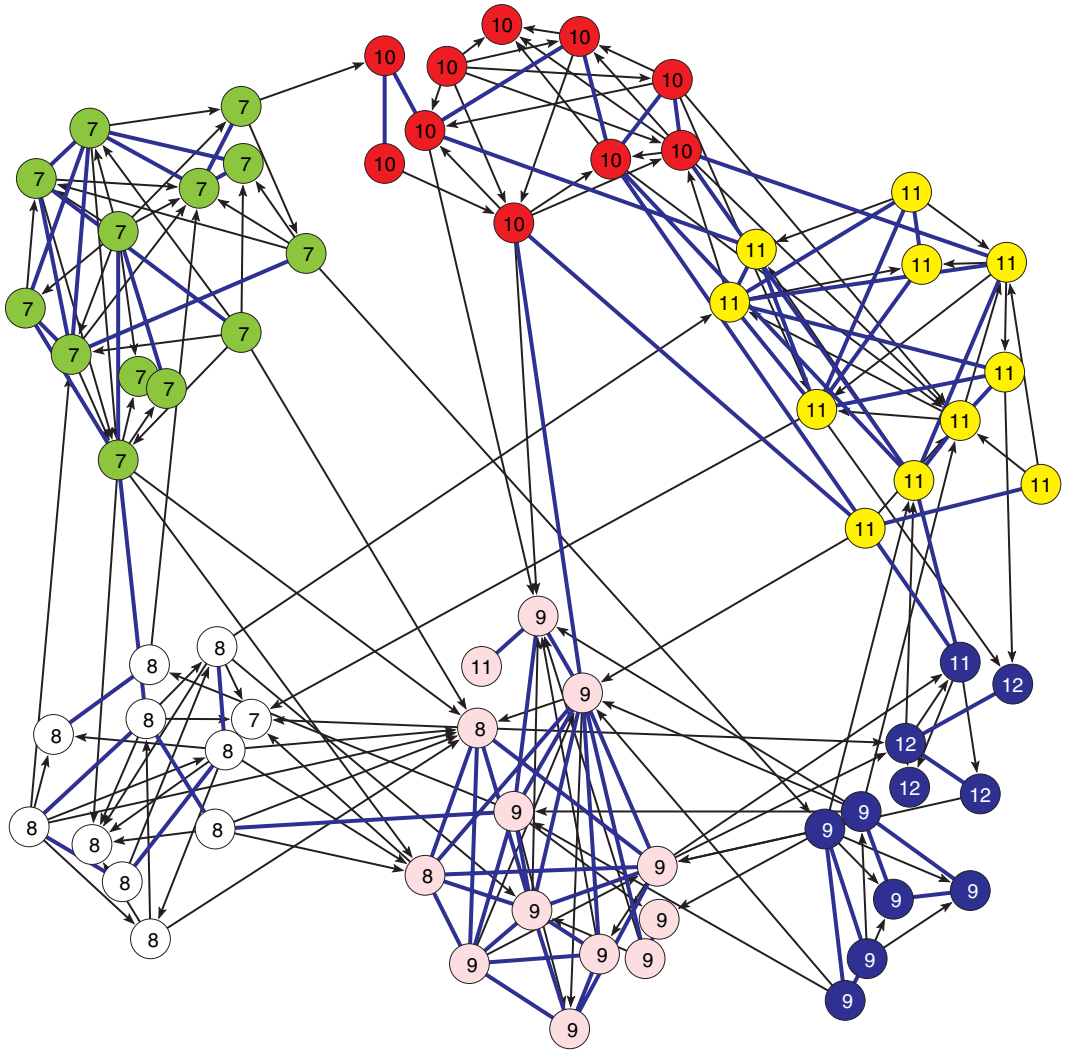

Fig. 13. Alternative partition of the adolescent health data

diagonal blocks are null. This specification has the unique partition that is shown in Fig. 13. We use grades to label vertices and undirected lines for reciprocated ties. The overlap between the grades and the clusters is shown in Table 5. There is consistency between their partition and ours. Both are readily interpretable.

Our approach is computationally simpler and also explicitly describes network structures. The appeal of the statistical approach includes the estimation of k and having an inferential foundation. It would be nice to couple these approaches to the benefit of both.

David Draper (*University of California, Santa Cruz*)

This excellent paper provides a nice example of contemporary likelihood and Bayesian analysis in an interesting social sciences setting; I have a comment about validation of the methods proposed. There are two main ways to evaluate the quality of a statistical method: *process* (do the assumptions on which the method is based seem reasonable?) and *outcome* (when you know what the right answer is, does the method tend to give you back known truth?). Of these two approaches, outcome evaluations are generally stronger than process assessments. For me, the methods of this paper pass a judgmental process test quite well (with one question; see the comment by Mendes and Draper). Regarding outcome, in their

Table 5. Partition clusters and grade levels

Grade	*Result 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	10	0	0	6
10	0	0	0	10	0	0
11	0	0	1	0	11	1
12	0	0	0	0	0	4

first example the authors seem happy when their Bayesian fitting method reproduces the cluster structure previously identified by the researcher who collected the data, and they mention in further support of the idea that the Bayesian result is 'good' that 'Overall the Bayesian estimate of the latent position cluster model produces greater distinctions between the groups . . .'. Of course these are not true outcome evaluations, because there is no comparison with known truth; in a sense they are more like another kind of process evaluation (do the results produced by the method seem reasonable?). The validation story appears a little stronger in the authors' second example, where the Bayesian fitting method by and large succeeds in inferring what grade the students were in without using that information in the fitting process; the authors are less happy to find that the maximum likelihood approach chose only two clusters, but this ignores the possibility that the dominant clustering is not by grade but by the (perhaps even stronger) distinction in the American schooling system between middle school and high school (indeed, Fig. 4 supports this two-cluster 'explanation', with the ninth graders occupying a transitional role between middle and high school; the authors say that they 'consider a single school of 71 adolescents from grades 7–12', but it is rare in the USA for all six of those grades to be taught in the same building). In the absence of known truth, both the two-cluster and the six-cluster solutions seem plausible, and plausibility (not validity) is the strongest conclusion one can claim. A more convincing validation exercise would involve

(a) finding a social network situation in which the actors perceive themselves as members of explicit social clusters,
(b) eliciting from each actor two kinds of information—the relational ties (e.g. 'I'm friends with persons a and g') and a form of personal 'truth' (e.g. answers to questions like 'I identify myself as a member of cluster X')—and
(c) trying to infer the personal truth from the relational tie information without using the former in the inferential process.

Has anyone tried this form of validation in the field of social networks?

Marijtje A. J. van Duijn (*University of Groningen*)
I congratulate the authors on proposing—and making available through accessible software—a very interesting network model that incorporates some important concepts from social network analysis.

The inclusion of transitivity (and reciprocity, in the case of directed relations) through latent positions is interesting, where cluster membership encompasses these (and possibly more) structural effects through the use of spatial proximity. In the two applications that were presented in the paper, the definition of space in two dimensions seems adequate. The authors, however, do not make any suggestion for the interpretation of these dimensions. It might be interesting to investigate whether these dimensions are related to other network or actor characteristics that are not included in the model, in the same way that the clusters in the applications were found to correspond—in varying degree—to known attributes. A logical next step would be to include these attributes in the model. I wonder about a possible trade-off between interpretability of clusters and model specification.

The random sender and receiver effects of the p_2-model (van Duijn *et al.* (2004), with accompanying software available at `http://stat.gamma.rug.nl/stocnet`) could be considered to define a latent space with a clear interpretation. Unlike the latent position cluster model and the earlier latent distance

models (Hoff *et al.*, 2002; Hoff, 2005; Shortreed *et al.*, 2006), the p_2-model uses the dyadic outcome as dependent variable and thus explicit incorporates the tendency of reciprocity within dyads. Its focus is on (fixed) actor and dyadic attribute (homophily) effects, and the model does not take into account transitivity or other triadic structural effects; nor does it consider the spatial representation of the network.

Model selection seems to be a topic requiring further investigation, in latent space models, and in the p_2-model (Zijlstra *et al.*, 2005). The somewhat heuristic Bayes information criterion approximation for the latent position cluster model seems to work quite well and is supported by the recent application of the Bayes information Monte Carlo criterion BICM (Raftery *et al.*, 2007). First results with BICM as a model selection criterion in the p_2-model are encouraging.

Katherine Faust and Miruna Petrescu-Prahova (*University of California, Irvine*)

Handcock, Raftery and Tantrum should be commended for presenting a principled basis for network scaling and node clustering. Our comments situate the latent position cluster model in relation to other social network analysis approaches and point to comparisons that facilitate interpretation.

The latent position cluster model contributes to a venerable tradition in social network analysis: combining spatial representation of social proximity with node clustering to identify subgroupings of actors. This combination often gives rich insight into social network structure, as seen in the authors' monastery and adolescent friendship examples. The latent position cluster model improves on extant methods by providing a principled way to determine dimensionality and number of clusters. It gives a model-based approach for network visualization with a precisely defined relationship between node distances and network ties. Combining clusters with positions shows internal differentiation within clusters and proximity of clusters relative to each other. These are valuable advances over many standard network methods for visualization and subgroup detection.

Regarding comparison of the two approaches, the authors observe that

> '... Bayesian estimate of the latent position cluster model produces greater distinctions between the groups than the two-stage estimate...'

(page 311). Indeed, clusters appear to be more easily distinguishable in Bayesian estimates. This is due to longer distances between cluster averages, but mostly to lower within-cluster variability in node positions, a point that is obscured by differently scaled axes in Figs 1 and 3. We refitted the two-dimensional, three-cluster model to the monastery data by using both approaches and display results in Fig. 14. Mean within-cluster distances for two-stage estimates are 1.164, 0.911 and 0.642, and for Bayesian estimates are

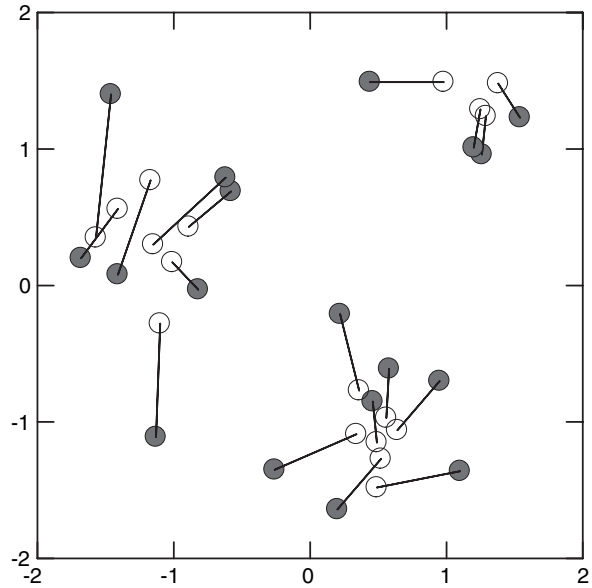

Fig. 14. Latent positions of monks in a monastery: two-stage maximum likelihood estimates (•) and Bayesian estimates (∘)

0.531, 0.317 and 0.279, for clusters 1, 2 and 3 respectively. Clearly, clusters from Bayesian estimates are more compact than are clusters from two-stage maximum likelihood estimates.

Assessing dimensionality is a valuable feature of the model that is not fully exploited in the paper. With regard to the monastery data, greater variability on the horizontal than vertical axis in Fig. 3, the apparent 'horseshoe' configuration, and only three arcs between clusters 1 (on the left) and 3 (on the right) suggest the possibility of a one-dimensional solution. The one-dimensional solution has BIC $= -360.6542$ compared with BIC $= -305.8171$ for two dimensions, and so it is not appropriate for these data.

Jonathan J. Forster (*University of Southampton*)

I have two questions concerning this interesting and stimulating paper. The main attraction of using the Bayes information criterion (BIC) in model comparison is that it can often be calculated by using outputs of standard packages. Given that the authors have already devoted considerable computational effort to carefully calculating or simulating posterior distributions, I wondered whether they had considered also more accurately approximating the marginal likelihood. It strikes me that the extra effort that is involved would be relatively little, and it would circumvent the difficulties in choosing a suitable n for the BIC formula (I do not find the argument for using the actual number of ties for the logistic regression BIC to be all that compelling). More generally, the paper considers spherically distributed clusters in a two-dimensional latent space. Could the authors give any insight into the benefits or problems that are associated with relaxing either of these assumptions?

Andrew Gelman (*Columbia University, New York*)

Social networks are important for their own sake and for their role in propagating phenomena such as political polarization. In a world full of disputes between and within nations, it is particularly important to have tools for studying the latent connectedness between people with disagreements and even hatreds, but who might be more tolerant of each other if they knew what connections they had in common.

I have little to add to the model or the statistical analysis except to point to the work of Watts *et al.* (2002), who noted that the social network is actually a union (i.e. overlapping) of networks from family, friends, church, work and so forth. Ideally, I think that a model of the social network would model these separate components. Along with this is the notion that networks evolve dynamically, with processes such as the completion of open triangles (if Ann knows Bob, and Bob knows Carl, then Ann is likely to meet Carl at some point); see, for example, Kossinets and Watts (2006). Perhaps the model of Handcock and his colleagues can be generalized to allow this time component (with the time points treated as latent data if they are not observed).

Finally, I encourage the researchers to think harder about how to present numbers such as those in Tables 1 and 3; for example, should we care that the estimate of β_0 under a particular model is '3.475'? For future work in this area, I recommend thinking carefully about what comparisons are of interest and then presenting the results graphically to learn about these comparisons (see Gelman *et al.* (2002)).

Steven M. Goodreau (*University of Washington, Seattle*)

The authors' work has many potentially important applications, of which two stand out for me. One is as an exploratory mechanism for understanding cases in which subpopulations that are defined by exogenous attributes are expected to form cohesive groups but do not. For example, in on-going work, colleagues and I are examining 59 of the school groups in the adolescent health study (the same study from which the authors draw a simple example) and are discovering that some forms of homophily and transitivity are both of universal importance. However, some groups display much more heterogeneity in their cohesiveness than others, most notably Hispanics and, to a lesser extent, native Americans and Asian Americans. We have considered several reasonable predictors for when these groups do or do not exhibit dyadic and triadic level cohesion, with limited success. The work that is presented in this paper could provide a novel method for distinguishing between the multiple ways that such groups can fail to be cohesive through analysis of both latent clusters and positions. Do they form more than one distinct subgroup, each of which is itself relatively cohesive? Or do different members of the population cluster tightly with other subgroups? Is there a large amount of uncertainty in cluster membership for some subset of actors? The latent position cluster model should allow us to distinguish between these possibilities in a statistically grounded way.

A second, and perhaps more widely applicable, use of these models could be as a tool for exploring goodness of fit. Social network modellers are often interested in knowing when their models have managed

to capture all of the relevant structure in a network, but traditional goodness-of-fit measures often do not provide a clear answer to this question. Although recent advances have been made in this area (e.g. Hunter *et al.* (2007)), such approaches necessarily require decisions about which particular features of network structure are important for a well fitting model to capture. The current approach would seem to provide an additional general method: adding the latent position cluster model as a 'residual' to a structural model to identify any remaining structure or clustering. Conceptually, such an approach would not only tell us whether such structure remained but also provide a sense of its nature. This would be an important addition to our toolkit for assessing model fit, which is so far an underexplored area of network analysis.

Priscilla E. Greenwood (*Arizona State University, Tempe*)

The beauty of the latent position cluster model is that no spatial setting is introduced. The setting is completely abstract, which means that the data set freely introduces its own structure via the inference step. The basic idea seems extremely flexible and can be applied with various regression models instead of model (2), and with other multivariate cluster–shape distributions replacing model (3).

The authors mention a possible epidemic interpretation. Suppose that a contagious disease runs its course in a community, and each individual tells us from whom he contracted the ailment. The method that is presented here can be used to infer the cluster structure of the epidemic. Epidemic parameters can be estimated simultaneously. This will be an interesting tool in spatial epidemic theory. Although epidemic data are rarely available in the form of directed ties, the method can be used as a simulation step. This example suggests two natural extensions of the latent position cluster model.

Suppose that we add a time structure to the model. Then, since the data give directed ties between nodes, we shall be able to infer a time ordering along the paths of the graph that is formed by these links, even though the directions in the data will not always be consistent. The result would give information about the evolutionary path of the epidemic through the community. Let us consider a genomic context. Inference about clusters together with time ordering, in a graph that is constructed from an alignment of homologous genes from several species, would produce a postulated phylogenetic structure.

A second natural extension would be to use the degree aspect of the data, the number of ties coming from each node, as an ingredient in the estimation of the number of clusters. In the epidemic context the degree data could be used for inference about the contagion parameter, either as a constant over the graph or locally within the clusters.

Katharina Gruenberg and Brian Francis (*Lancaster University*)

We congratulate the authors for having written such an inspiring paper. We would, however, like to point out two possible extensions which may arise out of real life data. The first relates to directed links. It is possible to measure links on a scale that allows positive as well as negative values for sociomatrices. Allowing for negative values allows the simultaneous modelling of 'like', 'dislike' and 'indifference'. For instance with the monk data we could investigate whether there are no ties between the 'Outcasts' and the 'Loyal Opposition' or whether their relationship is in fact one of possible mutual dislike. Alternatively, the same model could be used to model the reciprocal of dislike. The second point refers to the nature of social networks—homophily does not always exist in networks. People may be attracted to each other if they exhibit opposing ideas—the idea that opposites attract. Such relationships might exist in prison. Adapting the model to allow both for 'opposites attracting' and 'similars attracting' is a new challenge.

Christian Hennig (*University College London*)

I congratulate the authors for this very stimulating paper. I would like to contribute some thoughts about the model assumptions.

(a) The authors discuss the transitivity that is imposed by their model, particularly by the triangle inequality which is assumed to hold in the latent social space, but they do not explain how to check this model assumption. One possibility could be to apply a parametric bootstrap, i.e. to simulate new data sets from the fitted model and to compare the bootstrap distribution of the number of ties in triads with its observed value.
(b) It could be useful in some situations to allow more general covariance matrices within clusters. This enables, for example, elongated clusters, which have the reasonable social interpretation of modelling a group as spreading between two extreme points, which is not captured by spherical clusters.

(c) It could be useful to include the so-called 'noise component' as mentioned in Fraley and Raftery (1998) in the cluster model, because individuals who do not belong to any cluster may be found in many social networks.

Peter D. Hoff (*University of Washington, Seattle*)
Latent variable models of social networks can be motivated in a natural way: for undirected data without covariates we might view the nodes of a social network as being exchangeable, so that

$$\{Y_{i,j}, i<j\} \stackrel{\mathrm{d}}{=} \{Y_{\pi i,\pi j}, i<j\}$$

for any permutation π of the node labels. Aldous (1985) has shown that all such data can be expressed as $Y_{i,j}=g(\mu, z_i, z_j, \varepsilon_{i,j})$, where g is symmetric in its second and third arguments. Thus, the variation in any exchangeable sociomatrix can be represented with node-specific latent variables $\{z_1, \ldots, z_n\}$ and pair-specific noise $\{\varepsilon_{i,j}\}$. Bayesian estimation for the stochastic blockmodel of Nowicki and Snijders (2001) and the latent position model of Hoff *et al.* (2002) can be viewed as special cases of this general latent variable model. In the former, the zs are latent classes and g maps pairs of classes to between-class interaction rates. In the latter, the zs are vectors and g involves the Euclidean distance between them.

The stochastic blockmodel and the latent position model represent extremes of simplicity and complexity: the stochastic blockmodel implies that, conditionally on the values of the zs, all nodes within a common class share the same distribution over relationships. In contrast, the standard latent position model gives a different distribution for every node. The latent position cluster model that was presented by Handcock, Raftery and Tantrum nicely fills a void between the two approaches: a set of similarly acting, well-connected nodes will be identified and represented as a tight cluster, whereas nodes with unique behaviours will not be forced into ill-fitting groups.

As described in the paper, latent position models represent homophily. But this homophily is confounded with stochastic equivalence: similar values of z_i and z_j imply that i and j are likely to have a tie (since $|z_i - z_j|$ is small) and also have similar relationships to other nodes (since $|z_i - z_k| \approx |z_j - z_k|$). This correspondence is often present in friendship networks, but absent in networks such as the World Wide Web, in which 'hubs' connect to similar groups of nodes but not to each other. To separate homophily from structural equivalence we might consider an 'eigenvalue decomposition' model as described by Hoff (2006), in which the probability of a link between i and j is related to the form $z_i' \Lambda z_j$. By allowing entries of Λ to be either positive or negative, such a model can exhibit structural equivalence with or without homophily.

David R. Hunter (*Penn State University, University Park*)
The paper by Handcock and his colleagues provides an interesting and important extension of the latent space model of Hoff *et al.* (2002). A different extension of this work—which may also be applied to the current paper—allows for more explicit modelling of local network features, such as transitivity, by using an exponential random-graph model (ERGM).

If the matrix $\mathbf{y}$ denotes the entire network (i.e. the collection of all $y_{i,j}$), then equation (2) implies that

$$P(\mathbf{Y}=\mathbf{y}) = \frac{\exp(\beta_0^{\mathrm{T}} \sum_{i,j} x_{i,j} y_{i,j}) \exp(\beta_1 \sum_{i,j} |z_i - z_j| y_{i,j})}{\kappa(\beta_0, \beta_1)}, \tag{11}$$

where $\kappa(\beta_0, \beta_1)$ is a normalizing constant.

To simplify the notation in equation (11), let

$$\begin{aligned} g(\mathbf{y}, X) &= \sum_{i,j} x_{i,j} y_{i,j}, \\ h(\mathbf{y}, Z) &= \sum_{i,j} |z_i - z_j| y_{i,j}. \end{aligned} \tag{12}$$

Conditionally on the latent positions Z, the resulting model,

$$P(\mathbf{Y}=\mathbf{y}) = \frac{\exp\{\beta_0^{\mathrm{T}}\, g(\mathbf{y}, X) + \beta_1\, h(\mathbf{y}, Z)\}}{\kappa(\beta_0, \beta_1)}, \tag{13}$$

is evidently a canonical exponential family (see, for example, Lehmann (1983)) of distributions parameterized by (β_0, β_1) with statistics $g(\mathbf{y}, X)$ and $h(\mathbf{y}, Z)$. Therefore, conditionally on Z, model (13) is an ERGM ('graph' here is a synonym for 'network'). Snijders (2002) and Robins *et al.* (2007a) give literature reviews of these models, which are also called p-star models in the literature.

Importantly, model (3) is still an ERGM, conditional on Z, if the vector $g(\mathbf{y}, X)$ of network statistics of interest is not of the form (12) that allows the likelihood function to factor nicely as in equation (11). The simplest such 'non-factoring' models were considered by Frank and Strauss (1986), in which $g(\mathbf{y}, X)$ contained terms such as the number of triangles in $\mathbf{y}$, $\Sigma_{i<j<k} y_{i,j} y_{j,k} y_{k,j}$. Much recent work in the social networks literature has focused on development of useful statistics $g(\mathbf{y}, X)$ for modelling real network data (Snijders *et al.*, 2006; Robins *et al.*, 2007b), as well as explaining why some statistics, such as the number of triangles, lead to ERGMs that fail miserably at modelling these data (Handcock, 2002, 2003).

Model (13) would give the modeller a powerful tool for exploring network structure: for instance, if the latent positions and cluster assignments of the nodes change dramatically on the introduction of a particular network statistic into the ERGM, this suggests that the statistic captures an important aspect of network structure. Yet estimating parameters in a model such as equation (13) is quite difficult when $g(\mathbf{y}, X)$ is not of the form (12). In principle, the two-stage maximum likelihood estimation of Handcock and his colleagues should work, though the second stage would rely on a stochastic algorithm that is based on Markov chain Monte Carlo simulations such as those described by Hunter and Handcock (2006) or Snijders (2002). The Bayesian scheme that is implemented here is promising, but establishing a reasonable prior for the ERGM parameter β_0 is difficult. Despite the remaining challenges, this paper is a real step forwards.

Dirk Husmeier and Chris Glasbey (*Biomathematics and Statistics Scotland, Edinburgh*)

The authors have contributed an intriguing and stimulating paper to the growing literature on the statistical analysis of network structures. The model also provides a tool for visualizing networks, beyond existing algorithms such as Cytoscape (Shannon *et al.*, 2003).

Networks are of burgeoning interest in many fields, not least in post-genomic biology (see, for example, Wang and Chen (2003) and Milo *et al.* (2004)). Some biological interaction networks violate the underlying model assumptions, though. For instance, transcription factors regulating sets of unconnected genes and non-directly interacting proteins bound by the same protein recognition modules both lead to a violation of the transitivity condition. For this reason, model diagnostics would be a welcome addition to the work. Given that the authors have proposed a probabilistic generative approach, the application of diagnostics such as Bayesian p-values should be straightforward.

The model proposed could, in principle, contribute to post-genomic data integration. Consider, for instance, a situation where protein interactions inferred from yeast two-hybrid experiments are complemented by ribonucleic acid concentrations from transcriptional profiling with microarrays. The model allows us to infer the intrinsic trade-off between these two noisy and disparate data sets via equation (2), by treating the ribonucleic acid profiles as covariates and weighting their influence against the protein interactions via the two hyperparameters β_0 and β_1.

The authors approach the inference problem in terms of a hierarchical Bayesian model, sampling parameters from the posterior distribution with a Gibbs and Metropolis-within-Gibbs scheme, which is sound. Less sound, however, is inference on the number of clusters. The marginalization of the likelihood is carried out with respect to the parameters, but not the latent variables (i.e. the Z_is). Also, the Bayes information criterion approximation in equation (10) is rather restrictive. The Bayes information criterion assumes that the posterior distribution is multivariate Gaussian, ignoring differences in the eigenvalues of the covariance matrix, and the approach is hence compromised to the extent that this assumption is violated. Although a full reversible jump Markov chain Monte Carlo scheme might be computationally prohibitive, variational methods, which are currently very popular in the machine learning community, would presumably provide a much better approximation to the integration and might therefore provide a promising avenue for future research.

David Krackhardt (*Carnegie Mellon University, Pittsburgh*)

First, I want to emphasize the importance of the problem that is addressed by this paper. Cleanly identifying clusters of actors in a social system on the basis of their social ties is an age old pursuit of generations of scholars, from sociologists and psychologists to mathematicians (e.g. Luce and Perry (1949) and Cartwright and Harary (1956)). UCINET (Borgatti *et al.*, 2002), which is the most commonly used package for analysis of network data, has 20 distinct methods for finding clusters or groups, each with a plethora of suboptions and choices of parameter which, depending on the data, may yield wildly differing results. This dizzying array of 'solutions' begs the central question: given the observed data, what is the right number of clusters and what is their composition? Using the Bayes factors approach to answer this critical question statistically is a major step forwards out of this intellectual morass.

The paper quickly leads me to ask a couple of extending questions. First, how sensitive is this procedure to violations of the assumption on independence of dyadic observations? We know that even moderate amounts of 'network autocorrelation' in the data can dramatically affect estimates of standard errors and concomitant inference tests in traditional analytic procedures (Krackhardt, 1988).

Second, should we rely on empirical demonstrations of the model to provide us with evidence that the procedure is uncovering the true, underlying group structure? The fact that the procedure recovers the same structure in the Sampson data as other prior analyses could be because the networks are so clearly clustered that it does not matter what hammer you use to pound the data; they will always reveal the same story. In the case of ties between adolescents, the fact that their method cleanly shows discrimination between grades is interesting, but does that mean it was more accurate? Suppose that the result had not fallen along grade lines. Would that mean that the method was not accurately assessing real underlying clusters? Or, would it mean instead that networks were clustering on some other criteria?

Both of these questions could be addressed with appropriate Monte Carlo simulations. The advantage of such simulations is that you have control over 'truth', and by adding precise, known, and yet complex structures of noise, we can directly assess how well the proposed Bayesian method recaptures this underlying truth. Such simulations would help us to delineate the boundary conditions within which their method is truly powerful.

Jouni Kuha and Anders Skrondal (*London School of Economics and Political Science*)

We would like to query the authors' decision to advocate a method of model selection which is incoherent with their Bayesian estimation approach. In Section 4, they propose choosing the number of clusters by using the Bayes information criterion (BIC) statistic, calculated conditionally on posterior estimates $\hat{Z}$ of the latent positions. This uses maximum likelihood rather than maximum posterior estimates of the parameters and implicit prior distributions which differ from the priors that are specified in Section 3.2. It is thus only the estimated positions $\hat{Z}$ which actually depend on the results of the Bayesian estimation.

Would different results be obtained from an approach which was more consistent with the specified Bayesian model? To examine this, we generated 18 latent positions Z (scaled to have unit root mean square, as in Section 2) from a three-cluster model with parameters equal to the posterior medians in Table 1. Conditionally on these Z, we calculated, for models of 1–4 clusters, two approximations of $2\log\{P(Z)\}$:

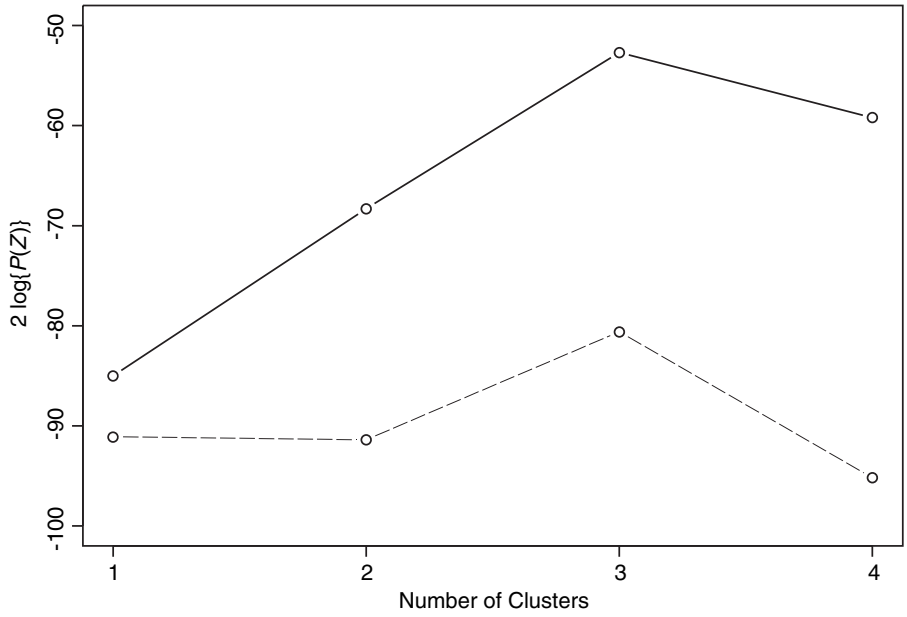

Fig. 15. Approximations of $2\log\{P(z)\}$ for models with 1–4 clusters (———, BIC statistic defined in Section 4; – – –, Laplace approximation): latent positions Z were generated from a three-cluster model with parameters specified by the posterior medians in Table 1 ($n = 18$)

the Laplace approximation (see equation (4) of Kass and Raftery (1995)) and the rougher BIC statistic of Section 4. The former depends on the posterior mode of the parameters and on the prior distributions that are actually used for the Bayesian estimation. For simplicity the values of Z were here the same for each model. In this case, $P(Y|Z)$ does not depend on the number of clusters, so model comparison is based on $P(Z)$ only.

Fig. 15 shows the values of the BIC and Laplace approximations of $2\log\{P(Z)\}$. Here both statistics correctly select the three-cluster model but there are some striking differences between them. The Laplace approximation imposes a larger penalty on the log-likelihood, so the prior distributions that are specified in Section 3.2 are actually less informative than the unit information priors that are used for the BIC. The effect of this is to increase the posterior probabilities of simpler models compared with the three-cluster model. In general, this means that a Bayesian model selection criterion based on the prior distributions of Section 3.2 may choose a model with fewer clusters than the BIC that is proposed in Section 4.

Having already used the Markov chain Monte Carlo machinery for Bayesian estimation, it would be natural to obtain direct estimates of Bayes factors as a by-product (without conditioning on $\hat{Z}$), instead of employing rough approximations of them. It seems plausible that a coherent approach that is based directly on Bayes factors would often favour smaller numbers of clusters than the approach which is considered in the paper.

Andrew Lawson (*University of South Carolina, Columbia*)

This paper is a very interesting example of the application of cluster modelling to a network domain. I have a few comments on the work.

First, the authors make a very strong parametric assumption about the latent positions in the social space in that

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

Here, the cluster form is forced to be symmetric, multivariate normal and around a mean level vector $\{\mu_g\}$. They also assume that the components are independent. In the subsequent model fitting, these assumptions appear to be unchallenged and yet it could easily be argued that there is a need for asymmetry and irregularity in social spaces. In other clustering applications this is not enforced (see for example Kim and Mallick (2002)). I am aware that the `mclust` software makes such assumptions, and so this affects the convenient implementation of the model. Have the authors considered relaxing these assumptions or examined the sensitivity of the model to these parametric restrictions?

Second, in the Bayesian model that is described in Section 3.2 the authors appear to fix certain parameters. For example, the β-parameters have fixed and relatively narrow variances, whereas in many Bayesian regression contexts these would have hyperpriors. This fixing of the hierarchy could lead to differences in estimates. Another example is the use of fixed variance priors for the mean parameters. Overall, this hierarchy truncation could be significant. Can the authors comment on the need for such truncation in their formulation? Does the implementation depend on this truncation?

Third, with regard to reversible jump and fixed G, the authors appear to avoid the idea of reversible jump sampling to allow for a dimension change in G. Indeed they do not even discuss the possibility. In addition, it is not clear from the paper whether G is fixed or allowed to vary. Clearly it would be feasible to assign a prior distribution for G and to sample it. Another possibility would be to formulate a binary variable selection model where

$$Z_i \sim \sum_{g=1}^{G} \psi_g \lambda_g \, \mathrm{MVN}_d(\mu_g, \sigma_g^2 I_d)$$

with ψ_g a Bernoulli selection variable which can be sampled. These choices appear to be simpler than the approach that is advocated by the authors.

Tim F. Liao (*University of Illinois, Urbana*)

I congratulate Handcock, Raftery and Tantrum for their contribution to the statistical literature on social network analysis. The model-based approach to analysing social network data represents a great leap forward: the method effectively overcomes the major disadvantage of previous methods where clique or cluster memberships are known, an assumption that is required by either the deterministic or the stochastic version of blockmodelling. I would like to focus on a potentially useful extension of the latent position cluster model, one that further relaxes the cluster membership assumption of the current method.

Cluster memberships can be defined as fuzzy rather than crisp and modelled as such. Two research traditions paved the foundation for this thinking. The sociological literature has long established that groups intersect within the person (Simmel, 1955), suggesting that one person can belong to multiple clusters, The idea was revisited in early social network analysis by Breiger (1974). In mathematics, Zadeh (1965) started a long line of research on fuzzy sets, with useful applications in engineering and in statistics (see, for example, Manton *et al.* (1994)).

The current method can easily extend to the use of grade of membership (or fuzzy membership) in estimating latent clusters. This can be achieved by defining uncertainty in cluster membership as a function of an actor's fuzzy membership, or $q_{ig} = f\{\mu_A(i)\}$, where $\mu_A(i)$ is the membership function of actor i in cluster A. Similar functions can be defined for clusters B, C, etc. Therefore, one actor may belong to multiple clusters to varying degrees. For the current examples, whereas the Sampson data may not need this extension (Fig. 3), the social network data from the National Longitudinal Study of Adolescent Health would more probably benefit from a fuzzy operation (Fig. 8).

As this should be a rather natural extension, I hope to see it developed in a sequel paper and implemented in a new version of `latentnet`.

Bruno Mendes and David Draper (*University of California, Santa Cruz*)

In their Bayesian fitting method the authors use conditional posterior model probabilities, which (as usual) are based on integrated likelihood values, and (as usual) integrated likelihoods can be highly sensitive to the manner in which diffuse prior distributions on the parameters of each model are specified (and this sensitivity can persist even with large sample sizes). If the authors had used a Laplace style $O(n^{-1})$ approximation to the logarithm of the integrated likelihood, they would have had to face this instability directly, because terms of the form $\ln\{p(\hat{\theta}_j|M_j)\}$ (where $\hat{\theta}_j$ is the maximum likelihood estimator or mode of the posterior distribution $p(\theta_j|y, M_j)$ for the parameter vector θ_j specific to model M_j; here y are the data) would arise in the Laplace approximation and could easily vary unstably as a function of the details of the diffuse prior specification. They appear to avoid this problem by using a cruder $O(1)$ approximation based on the Bayes information criterion, in which prior specification details are swept under the rug. The something-for-nothing bell is ringing in the background here: apparently one can get around a fundamental difficulty (which does not necessarily go away as the amount of data increases) with integrated likelihoods just by adopting a cruder approximation to them. Perhaps the authors can clarify.

Gesine Reinert (*University of Oxford*)

The authors are to be congratulated on their paper; it provides a novel approach linking statistics and social network analysis.

An interesting tangent is the recent statistical physics approach to networks. The most basic such construction is the *Watts–Strogatz model*, where random shortcuts are added to a fixed lattice, the end points of the shortcuts being chosen uniformly. Slightly more complicated models arise in network growth models, where a new vertex creates a (fixed or random) number of links to existing vertices, with possibly preferential attachment rules. These classes of so-called *small world networks* are claimed to provide suitable models for social networks such as scientific collaboration networks and Internet dating networks; for an overview see for example Dorogovtsev and Mendes (2003).

As customary in statistical physics, many of the small world network results are of asymptotic nature. In contrast, the problems that were studied by the authors involve only a small number of vertices, so asymptotic regimes may not be of any direct interest. In addition, small world network models may not capture some of the important features in the data. Yet even asymptotic small world network results could potentially be relevant, not only because increasingly more large social network data sets become available, but also because such results help to understand better the qualitative behaviour of networks. An example are results on the emergence of a giant cluster (see Durrett (2006) for an introduction and Bollobas *et al.* (2006) for recent progress), which could relate to percolation-based clustering algorithms as in Sasik *et al.* (2001).

Beyond Bernoulli random graphs, for all these network models assessing model fit remains an open question. Often networks are summarized by using the clustering coefficient, the average shortest path length, the average vertex degrees or the number of occurrences of certain network motifs. Ideally (at least asymptotic) distributions for some summary statistics would be available to derive parameter estimates and to develop rigorous statistical tests for model fit. For Watts–Strogatz small world networks a few results can be found in Barbour and Reinert (2001, 2006), but much more research on these issues is needed.

Although networks have recently received considerable attention not only from social scientists but also from statistical physicists, few statisticians have taken up the challenge of contributing to this field. The paper under discussion might help to reverse this trend.

Sylvia Richardson and Alex Lewin (*Imperial College London*)

The authors are to be congratulated on their stimulating paper that will foster the application of the same ideas in many different domains.

Our comments relate to two aspects.

(a) The authors hardly discuss the choice of the dimension d of the latent space and their examples always use $d=2$. It would be reasonable to expect a close relationship between d and the number of cluster, a relationship that is not discussed. For a large complex network, if the dimension chosen is too low, then possibly more clusters will seem to be necessary. Could d be included as a parameter in the analysis so that joint inference is made on d and the number of clusters?

(b) It is unclear to us whether the additive formulation (2) is the best way of including homophily on observed attributes. There is an interplay between the effect of covariates on the log-odds for links and the latent space which captures the effect of hidden characteristics of the social actors. A useful parallel can be drawn with ecological regression and spatial patterns of disease. Usually the ecological regression equation for relating the underlying log-relative-risk for a disease in an area i to area-specific covariates X_i is written as

$$\log(\theta_i) = \beta_0^{\mathrm{T}} X_i + s_i$$

where s_i is a Markov random field that captures *residual latent spatial structure* that is not accounted for by the covariates. This assumes no interaction between the covariates and the spatial space, and when this is not reasonable an interaction with space is considered instead, i.e. β_0 becomes indexed by i (see for example Gelfand *et al.* (2003)). For social networks, we feel that such interaction is likely and, hence, including covariates solely as a fixed effect might not be appropriate. For example, you might expect girls and boys to mix more in older years and therefore the influence of gender similarity to be different for each age cluster. Thus a more realistic model would investigate whether the homophilic effect of the covariates differs for different clusters. A useful extension of equation (2) might thus be

$$\text{log-odds}_{ij} = \beta_{\delta_i} X_{ij}\, I(\delta_i = \delta_j) + \beta_0 X_{ij}\, I(\delta_i \neq \delta_j) - \gamma |z_i - z_j|$$

where δ_i is the allocation label in the clustering for actor i. We believe that such an extension would enhance the capacity of the model to account for complex network structure and we would welcome the authors' thoughts on the interplay between covariates and the cluster structure.

D. M. Titterington (*University of Glasgow*)

I have two comments about this interesting paper.

If we denote the set of all group memberships by K and let ϕ denote all parameters, then a key factorization is

$$P(Y, Z, K|X, \phi) = P(Y|Z, K, X, \phi)\ P(Z|K, X, \phi)\ P(K|X, \phi). \tag{14}$$

In the paper a specialized version of this is used, corresponding to

$$P(Y, Z, K|X, \phi) = P(Y|Z, X, \phi)\ P(Z|K, \phi)\ P(K|\phi). \tag{15}$$

Furthermore, if there is no covariate information X, as is the case in both main examples, then this becomes

$$P(Y, Z, K|\phi) = P(Y|Z, \phi)\ P(Z|K, \phi)\ P(K|\phi). \tag{16}$$

The Bayesian calculations in Section 3.2 essentially use the formula in equation (15), together with a prior $P(\phi)$, as a basis for estimating $P(Z, K, \phi|Y, X)$ by using Markov chain Monte Carlo sampling. In contrast, in the first stage of the two-stage method in Section 3.1, the method of Hoff *et al.* (2002) takes the first factor on the right-hand side of equation (15), namely $P(Y|Z, X, \beta)$, where β is the part of ϕ corresponding to that factor, and maximizes it with respect to Z and β, with the resulting $\hat{Z}$ being referred to as the 'maximum likelihood' estimates of the latent positions. My first comment is to indicate some anxiety over this, because it is well known that treating 'missing values', such as latent scores, as 'parameters' in this

way can lead to problems such as biases in the estimators of the genuine parameters; see for example Little and Rubin (1983) and Marriott (1975). However, I concede that the normative maximum likelihood approach would be computationally difficult. It would involve using the EM algorithm to estimate ϕ, with complete-data likelihood given by whichever of equations (14), (15) and (16) is appropriate, and then obtaining values for the latent positions Z in the same spirit as the calculation of factor scores in factor analysis.

This brings me to my second point. Suppose, in contrast with equation (15), we factorize $P(Y, Z, K|\phi)$ as

$$P(Y, Z, K|\phi) = P(Y|Z, K, \phi)\, P(Z)\, P(K|\phi).$$

If the variables in Y are continuous and if $P(Z)$ corresponds to $Z \sim N(0, I)$, then this gives the mixture of factor analysers model for Y; see Ghahramani and Beal (2000) and Fokoué and Titterington (2003), and also Fokoué (2005) for a version that incorporates X-like covariates. It would be appropriate to describe the version with binary variables in Y as a mixture of latent trait models (Bartholomew, 1987) or a mixture of density networks (MacKay, 1995). I wonder whether this variation would produce interesting results in the contexts that are covered by the paper. I suspect probably not, at least so far as interpretability is concerned, but the relationship between the two types of model may be of interest.

Stanley Wasserman (*Indiana University, Bloomington*)

At a reception about 10 years ago, part of a memorial tribute to Cliff Clogg at Penn State University, a well-known, and very good, statistician–sociological methodologist chatted with me about social network analysis. I was surprised that this well-known person, after telling him that I did network analysis, had the view that network analysis was just a bunch of indices, with little thought given to statistical models. And, I felt the weight of his accusation. He was incorrect, of course, but it was a common misperception at that time.

There is no question that network analysis has come a long way over the past decade, spurred on to some extent by the many researchers doing statistical modelling and ordinary people who are interested in networks. Here in the States, there is even a new television show on the ABC network named 'Six degrees'. With networks pervasive in our 21st-century popular culture, it is pleasing to know that we are learning what to do with network data. The paper under discussion here, by Professor Raftery, Professor Handcock and Dr Tantrum, is a very fine piece of mathematics, a perfect example of the growth of the discipline. It certainly advances network science, but it leaves me with a few questions.

First, what will be the fate of this clever model? Will it be ignored, as Hoff *et al.* (2002) has been? Many statistical approaches to networks (such as correspondence analysis and stochastic blockmodels (Wang and Wong, 1987)—which include the models that are described here) have been little used. Could a mere mortal, a social networker from, say, social work, fit this model? A friend of mine years ago remarked that network data are more complicated than the models that are used to study them. I think that the opposite is now true.

Second, what has happened to network *data analysis*, as opposed to statistical network modelling? Sure, we have great models and the computing ability to fit them by using appropriate and correct estimation techniques, but very little thought has gone into questions such as 'why this model and not that one?'. How does model A compare with models B–Z on a wide range of data sets? The authors ignore this issue. It may take years to answer questions such as these. We are just now making progress on understanding the exponential family p^* (using good ideas such as those in Goodreau (2007)), but we need more data-oriented papers such as Holland and Leinhardt (1981).

Network research needs more of Cliff Clogg, a good sociologist and a superb statistician, who cared about data, and less of Bayesian formalism.

The moral of my story at the beginning of my comment is that this very person is now doing research on excellent, and sophisticated, network models. I feel partially vindicated!

Adriano Velasque Werhli (*Biomathematics and Statistics Scotland, Edinburgh*) **and Peter Ghazal** (*Scottish Centre for Genomic Technology and Informatics, Edinburgh*)

Although the title of the paper suggests that the method proposed is restricted to the analysis of social networks, it is interesting to investigate whether it has the scope for wider applications to biomolecular interaction networks. For this we have applied the algorithm to a genetic network that is related to the action of interferons, which play a pivotal role in modulating the innate and adaptive mammalian immune system. The network is shown in Fig. 16(c). We have applied the algorithm in the same way as described in the paper, applying standard diagnostics to test for convergence of the Markov chain Monte Carlo simu-

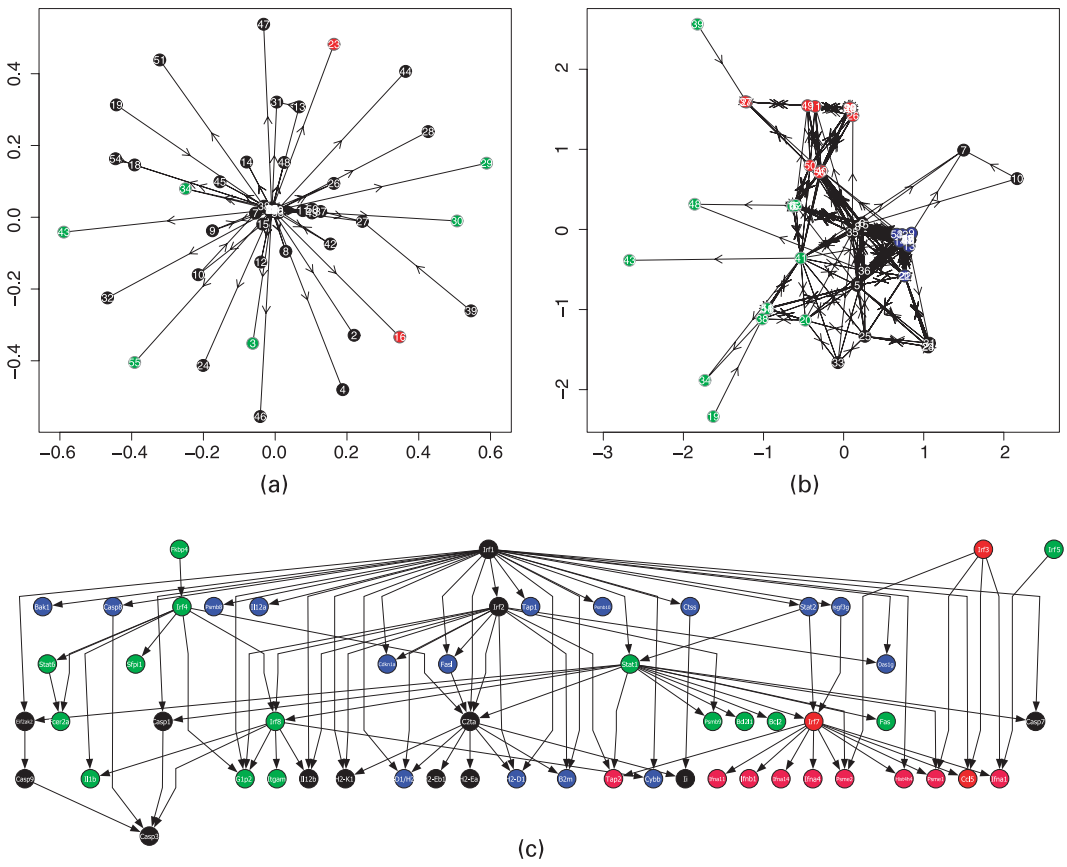

Fig. 16. Genetic network related to the action of interferons: (a) positions of the nodes in the latent space; (b) resulting positions of the nodes in the latent space after application of the authors' modification; (c) the genetic network

lations. Fig. 16(a) shows the positions of the nodes in the latent space, obtained for the number of clusters with the highest marginal likelihood score (ngroups = 3). It is obvious that no clear cluster formation is found, and the cluster assignment that was predicted was not biologically meaningful.

The reason for the failure of the algorithm becomes clearer when investigating the interferon gamma pathway more closely. There are various hub nodes connected to sets of peripheral nodes that are not themselves interconnected, and this violates the transitivity assumption on which the algorithm is based. To put this to an empirical test, we modified the interferon network as follows. We identified seven central regulators (i.e. hub nodes): Statl, Irfl, Irf7, C2ta, Irf3, Irf2 and Irf4. For each regulator, we completely interconnected all the regulated genes with bidirectional edges and, in addition, introduced bidirectional edges between the regulators and regulatees. This is to ensure the formation of clique structures that satisfy the transitivity condition. We then applied the method proposed to the modified network. The resulting positions of the nodes in the latent space are shown in Fig. 16(b). Fig. 16(c) shows the original network, where the shading of the nodes indicates the cluster membership (again, we used the number of clusters that maximizes the marginal likelihood: ngroups = 4). The cluster formations are now much more distinct and are clearly related to the regulators and their regulated genes. (There is no perfect agreement owing to interconnections between the cliques and violations of the transitivity condition in other parts of the network).

This analysis indicates that the algorithm proposed is not generalizable to molecular biological interaction networks that inherently violate the transitivity condition.

The **authors** replied later, in writing, as follows.

We thank all the discussants for their stimulating comments. The large number and wide range of discussions suggest that the statistical analysis of social networks is a developing area that is poised for rapid growth. Many potential applications were mentioned, including to epidemics (Greenwood), post-genomic data (Husmeier and Glasbey), biomolecular interaction networks (Werhli and Ghazal) and rank data (Gormley and Murphy).

We appreciate the many positive comments about the latent position cluster model. In particular, we would underline Snijders's comment that latent structure models allow data that are missing at random to be handled almost trivially. This point was not made in our paper, and it is important because often much of the data about a network of interest is based on network sampling or subject to out-of-design missingness.

Social network characteristics

Our model was designed to take account of homophily on observed attributes, transitivity and clustering, but it did not incorporate other important features. One of these is what Snijders calls prominence and is also referred to as activity, sociability or popularity, namely the fact that some actors tend to send and/or receive more links than others, sometimes by a large margin. Greenwood emphasizes the importance of this for applications to infectious disease epidemics. Note that in our examples this was not an important feature of the data, as can be seen from Figs 1 and 4, for example. This was in part because the data collection method discouraged it; for example, the school students in the adolescent health data set were invited to name no more than 10 friends, and most did name close to that number, so the tendency to send links did not vary greatly between students.

It seems most natural to allow for this by adding random sender and receiver effects to equation (2) of our paper; this would be a small technical modification to the model. This is similar to the specification of random effects in the p_2-model by van Duijn *et al.* (2004), as pointed out by Snijders and van Duijn. Hoff suggests an eigenvalue decomposition model (Hoff, 2007) as an alternative. This seems less easily interpretable than a random-effects model but allows the separation of homophily and structural equivalence. In our experience, it is also computationally efficient, which is important for scaling the methods to larger networks.

Another important feature that our model does not include is what Snijders calls hierarchy and Butts calls asymmetry, namely the tendency in a given dyad for one member to send links and the other to receive them. As Snijders points out, this can also be represented at least partly by sender and receiver random effects, Butts suggests a simple and elegant generalization of this idea, in which each actor has a latent, possibly vector, 'vertex potential'. This is an important contribution. Overall, we agree with Snijders's view that activity and popularity dimensions should be included by default in latent space modelling of social networks, and this is easy to do in our modelling framework.

Breiger, and Gruenberg and Francis point out that some social networks exhibit negative affect, as a result of which people with opposite attributes tend to attract. The clearest example is that of heterosexual sex networks. When the relevant attributes are observed, as in the sex network case, this could be dealt with naturally in our model by the $\beta_0^{\mathrm{T}} x_{i,j}$-term in equation (2), where $x_{i,j}$ represents dissimilarity (e.g. being of the opposite sex) and β_0 is positive. It is indeed a challenge to adapt the model to the situation where opposites attract and the relevant attributes are unobserved, as Gruenberg and Francis remark.

Lawrance and Krackhardt ask how sensitive our results are to the conditional independence assumption in equation (1). We think that the answer is 'not very'. In our model, links are conditionally independent given the unobserved latent variables z_i; thus unconditionally they can be highly dependent. Indeed Hoff, citing Aldous (1985), points out that all social network data of this type can be represented as conditionally independent given some actor-specific latent variables, which provides some theoretical basis for thinking that the conditional independence assumption is not restrictive. As noted by Hunter and Snijders, this assumption can be tested by incorporating a more general exponential random-graph model (ERGM).

Model-based clustering specification

Our model specifies the distribution of latent positions within a cluster to be multivariate normal with a spherical covariance matrix. Robinson, Forster, Hennig and Lawson ask whether we could relax this assumption to allow a more general, non-spherical covariance matrix. We did experiment extensively with such a model and found that the results were often unstable and difficult to interpret. This seems to be because the amount of information in the data that is used to define, say, a cluster of seven monks is actually quite small, consisting of a small number of binary observations, and is not enough to specify a general covariance matrix with adequate precision. With the simpler model that we used, we did obtain stable and interpretable results.

Nevertheless, it is possible that in some cases a non-spherical model could be useful. For example, Bearman *et al.* (2004) reported 'chaining' effects in romantic networks of adolescents, and it is possible that such clusters could be represented by long thin mixture model components, with covariance matrices that have a high ratio of largest to smallest eigenvalues.

Robinson suggests the use of non-Gaussian components in the mixture model, and Atkinson and Longford point out that a mixture of normal distributions can represent a non-Gaussian shape rather than clustering. Our experience, however, is that network data do not provide enough information to support the use of non-Gaussian components or to lead to the use of more than one Gaussian component for a single cluster. These issues can be important for clustering observed data, but they seem much less relevant when clustering latent positions that are not very precisely determined by the data.

Hennig suggests the addition of a low intensity uniform noise component to the mixture model (3) to represent isolated actors with few or no links. This is an excellent idea, as isolated actors are common in social networks and are difficult to model. The use of a low intensity uniform noise component to represent outliers in model-based clustering was proposed by Banfield and Raftery (1993), and Hennig (2004) has shown that it leads to methods with good classical robustness properties when applied to observed data.

Choice of distance

We used the Euclidean distance between latent positions to specify our model. This has the advantage that the resulting positions can be represented in Euclidean space, which is useful for visualization and interpretation. However, as Snijders points out, other distances, such as the ultrametric, could also be used.

Breiger and Gelman point out that in practice people belong to multiple networks, and that our model does not account for this. One way to do so could be to change the dissimilarity measure in the latent space model. For example, we could replace the Euclidean distance $|z_i - z_j|$ in equation (2) by the co-ordinatewise minimum dissimilarity, $\min_k |z_{ik} - z_{jk}|$, where the z_{ik} are the co-ordinates of z_i. This could be interpreted as follows. If each co-ordinate corresponds to a different component network (family, friends, work, neighbourhood, etc.), then proximity on any one of them will be enough to make the chance of a link high. For example, if Bob works in the same office as Carl but lives far from him, they are almost as likely to form a link as if they lived closer. Each network could be specified by more than one co-ordinate. This should not make the estimation problem more complex.

Extensions

Besag, Breiger and Gelman point out that networks evolve dynamically and that extending the model to incorporate this would be useful. Greenwood points out that this is particularly important for modelling infectious disease epidemics. We agree, and a start has been made on this by Westveld and Hoff (2005).

Another important extension is to the case where links are not binary, but quantitative, e.g. a measure of how friendly Bob and Carl are rather than just whether or not they are friends. Such a measure could be continuous valued, or a count, such as the number of times that they meet per month. It could also be categorical, for example, if the link could be negative (e.g. dislike) or positive, as pointed out by Gruenberg and Francis. Such an extension can be readily accommodated within our framework, by replacing the logistic regression of our equation (2) by another response model. This could be a generalized linear model, as proposed by Hoff (2003). See also Oh and Raftery (2003).

Snijders and Hunter suggest combining the latent position cluster model with the ERGM class. This would allow the direct representation of structural signatures hypothesized under social theories (e.g. triadic balance). The interpretation of the latent component of the model then changes as it then represents residual social structure. Steps in this direction have been made by Handcock *et al.* (2003b). One way in which this combination would be immediately useful is suggested by van Duijn's discussion. Our model does not fully model reciprocity. This could be done by replacing our equation (2) by a bivariate binary response model for $y_{i,j}$ and $y_{j,i}$ jointly, where the model is a p_2-model as specified by van Duijn *et al.* (2004), but specified conditionally on the distance $|z_i - z_j|$.

Robinson, and Richardson and Lewin raise the issue of how the dependence on observed attributes and the clustering should be jointly specified. This is an important and unresolved issue that we did not investigate in our paper beyond writing down equation (2). Social network researchers would refer to Richardson and Lewin's extension as differential homophily by cluster, and it is an interesting possibility. More basically, the question of whether dependence on observed attributes is best represented by our equation (2) at all is an open question. An alternative would be to allow the observed attributes to influence group membership probability, leading to a mixture-of-experts model, as suggested by Gormley and Murphy.

Titterington suggests a different factorization of the likelihood for our model, which suggests possible alternative models. He suspects that this would not produce interesting results in our context, and we must agree, but his discussion still places the modelling in a broader framework that could be productive.

Alternative models

Perhaps the most influential alternative to statistical estimation of probability models for social networks consists of models from statistical physics, such as small world networks and models that are based solely on degree distributions, as noted by Reinert; see Newman (2003). We must agree with her that these models may not capture important features of the data, but that some of the results from this literature may be helpful. We also agree that the physics literature lags in model fitting and assessment, and we note that statisticians have started to contribute here (Handcock and Jones, 2004; Handcock and Morris, 2007).

The generalized blockmodels that were discussed by Breiger and Doreian, Batagelj and Ferligoj are based on deterministic algorithms, and as such do not provide a statistical basis for estimation, inference and choice of the number of groups. Nevertheless, these results may give insight into network structure that could be useful in statistical modelling, and so we welcome the suggestion to couple the two approaches. We note that our model is not simply a stochastic blockmodel, as the actors are not structurally equivalent given the cluster membership, but are, given the latent positions. Thus members of the same cluster are heterogeneous (e.g. Victor and Romuald in the Loyal Opposition).

Airoldi, Blei and Fienberg, and Liao discuss the possible use of the grade of membership model, either in place of or in combination with the mixture model that we use here. This would allow actors to belong to several groups with different 'grades of membership'. As Liao notes, the idea that individuals are defined via their group memberships goes back a century to the work of Simmel. These models would indeed be appropriate when the objective is to represent identity as a function of latent group memberships.

Sweeting suggested using a Dirichlet process mixture or similar model for the model-based clustering component of our model. This seems well worth investigating. Our experience with the Dirichlet process mixture model suggests that care is needed, however; see, for example, Petrone and Raftery (1997). For example, conditionally on a given number of groups, the Dirichlet process mixture prior tends strongly to favour very unbalanced groups, which may not be appropriate.

Model fit

We did not report much assessment of the fit of the model in absolute terms in our paper, and it is indeed important to do this. Given our estimation method, the most natural framework for this is that of posterior predictive checking (Gelman *et al.*, 1996), as alluded to by Husmeier and Glasbey. The statistics that are used for this could be descriptive network measures capturing important aspects that we want to reproduce; Reinert lists some common statistics, and Snijders suggests some new measures that could be used. As noted by Goodreau, a general framework for goodness of fit has been developed by Hunter *et al.* (2007) and Goodreau (2006), and posterior predictive checks are implemented in the software Handcock *et al.* (2004).

Hennig suggested using a parametric bootstrap, and this could be viewed as an approximation to posterior predictive checking. Snijders noted that this would be time consuming, however, and one advantage of posterior predictive checking here is that it could follow from our Bayesian estimation method with modest computational effort.

Reinert suggests using asymptotic theory to obtain the distributions of test statistics, whereas Goodreau suggests adding a latent space model as a residual to see whether there is any remaining structure. These suggestions seem worth pursuing. Goodreau's suggestion should be particularly helpful in decomposing the variation due to observed covariates, structural signatures (via ERGM terms) and residual social structure.

Draper suggests that we use ground truth to assess the model, in particular the clustering. What we did in the monks example is similar to what he suggests. The 'known' clustering that we used there as ground truth was based on a large amount of information, including ethnographic study by the researcher S. F. Sampson, who lived in the monastery for a year and observed the monks' interactions closely. We know of no formal attempts at outcome validation using personal 'truth' but note that self-identification may not be definitive. Lawrance's suggestion that we validate the model by applying it to academic departments could give an even more compelling form of ground truth!

Model choice

We used Bayes factors, approximated by a version of the Bayes information criterion (BIC), to compare models. Krackhardt pointed out that there are dozens of competing methods for finding clusters in social

network data that can give wildly differing results, but until now there has been no clear way to choose the best method. We agree strongly with him that using Bayes factors is a 'major step forwards out of this intellectual morass'.

Forster, and Kuha and Skrondal suggest using an integrated likelihood from the Markov chain Monte Carlo (MCMC) output rather than our BIC approximation. Although it would seem that this should be easy, it has turned out to be surprisingly difficult to find a generic method for doing this. Raftery *et al.* (2007) reviewed this literature and proposed a criterion called BICM based on the MCMC output. We are glad that van Duijn could report favourable results with this criterion; Gormley and Murphy (2007) also reported good results with BICM in a different latent space model.

Husmeier and Glasbey say that we should have integrated out the latent positions when computing the Bayes factors, but Snijders found that what we did, i.e. keeping the latent positions fixed in the Bayes factor calculations, was reasonable. This is clearly debatable, but our argument for doing what we did seems to have been acceptable to most discussants. Husmeier and Glasbey also assert incorrectly that the derivation of the BIC assumes that the posterior distribution is multivariate normal with an isotropic diagonal covariance matrix, but in fact the result that the BIC provides an $o(1)$ approximation is valid in much greater generality (Kass and Wasserman, 1995).

Kuha and Skrondal, and Mendes and Draper suggest using the Laplace method to integrate out the parameters, rather than the BIC (while keeping the latent positions fixed). This seems like a good idea, especially since the BIC is derived from the Laplace method. However, Mendes and Draper point out that the sensitivity of model choice to the prior would then be an issue. Kuha and Skrondal reported some assessment of that sensitivity and found that with our priors the resulting integrated likelihood tends to indicate less evidence than the BIC for more complex models. This indicates that our prior is in some sense more spread out than the unit information prior that underlies the BIC.

This raises an interesting point. Typically, Bayesian estimation is not much affected by making the prior flatter, but Bayesian model choice can be. Our priors were designed for estimation, so we made sure that they were at least as spread out as reasonable prior information, but we did not devote much effort to making sure that they were not too spread out, which is also necessary in the model choice situation. If the priors were to be used for computing Bayes factors more exactly, we might need to revisit them to ensure that they are not too spread out. Overall, we feel that our BIC approximation provides a reasonably simple and robust method for model comparison in the present context.

Choice of dimension

We used two dimensions for the latent space throughout, but it would be possible, and perhaps desirable, to make the choice of dimension data dependent. Oh and Raftery (2003) showed how to do this by using Bayes factors for a similar model, based on Oh and Raftery (2001). They found, perhaps surprisingly, that there was little interaction between the dimension of the latent space and the number of clusters. If this also holds in the present context, there would be little need for the simultaneous choice of dimension and number of clusters that was suggested by Richardson and Lewin. Note that a direct use of the BIC for choice of dimension, as done by Faust and Petrescu-Prahova, is not correct for choice of dimension here.

Raftery *et al.* (2007) applied their methods for estimating integrated likelihoods from MCMC output to precisely this problem and found that for the monks network the choice of two dimensions was favoured. This could provide a simpler and more generic solution to the problem. For the adolescent health data, a third dimension does not lead to clear separation of the higher grades. The choice of one dimension does lead to groups approximately ordered in grade, although with substantially less definition than the two-dimensional version.

Estimation

Leslie suggested improving the efficiency of our MCMC algorithm by updating the group memberships and the latent positions simultaneously. This sounds like a good idea, although it is an empirical question whether the gain in efficiency is worth the resulting greater complexity of the algorithm. Kent points out that it is the shape of the configuration of latent positions that is important. This is correct, and we took account of that by the Procrustes step in our algorithm. However, we would welcome further insights from shape analysis. Snijders recommended assessing how well the data determine the latent positions and suggested sensitivity analysis for this. In fact, the posterior distribution of the latent positions (after the Procrustes step) gives an assessment of this that comes right out of our method, although we did not have space to show it in the paper. Sophisticated plotting of the posterior is implemented in the package, and example code is given to produce the plots for the monks data.

Besag, and Blei and Fienberg asked about the important issue of scalability of the algorithm to larger networks. We have successfully applied the methods to networks with up to 3000 nodes. If necessary, for very large networks, it may be possible to approximate the algorithm without compromising its essential features by case–control sampling of ties in the computation of the likelihood or something like the ICM algorithm (Besag, 1986).

Incidentally, Besag asked why we showed results for only one school in the adolescent health data. This was convenient for presentation, but we have applied the methods to all the schools (Hunter *et al.*, 2006; Goodreau, 2007). We found that the method provides insight into both clusters and segregation. However, for medium-to-large schools (above 500 students) the visualization methods need to be more sophisticated to extract the information (e.g. zooming and slicing).

General

Wasserman's frustration in understanding recent advances in statistical network modelling is understandable. Making complex models accessible to practitioners is important. We believe that providing high quality software is becoming an essential part of publication, not least because it allows others to evaluate and critique the models proposed (Handcock *et al.*, 2003a, 2004; Boer *et al.*, 2003). Using this software, social network practitioners, including those from social work, routinely fit these and ERGMs in a class that is taught by one of us (Handcock)! This is part of the reason that the model of Hoff *et al*. (2002) is much used and extended (as the discussants testify).

References in the discussion

Airoldi, E. M. (2006) Bayesian mixed membership models of complex and evolving networks. *Doctoral Dissertation*. School of Computer Science, Carnegie Mellon University, Pittsburgh.

Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2007a) Admixtures of latent blocks with application to protein interaction networks. To be published.

Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2007b) Combining stochastic block models and mixed membership for statistical network analysis. *Lect. Notes Comput. Sci.*, to be published.

Aldous, D. J. (1985) Exchangeability and related topics. *Lect. Notes Math.*, **1117**, 1–198.

Atkinson, A. C. and Riani, M. (2007) Exploratory tools for clustering multivariate data. *Computnl Statist. Data Anal.*, to be published (doi:10.1016/j.csda2006.12.034).

Atkinson, A. C., Riani, M. and Cerioli, A. (2004) *Exploring Multivariate Data with the Forward Search*. New York: Springer.

Atkinson, A. C., Riani, M. and Cerioli, A. (2006a) An econometric application of the forward search in clustering: robustness and graphics. In *Prague Stochastics 2006* (eds M. Husková and M. Janžura), pp. 63–72. Prague: Matfyz.

Atkinson, A. C., Riani, M. and Cerioli, A. (2006b) Random start forward searches with envelopes for detecting clusters in multivariate data. In *Data Analysis, Classification and the Forward Search* (eds S. Zani, A. Cerioli M. Riani and M. Vichi), pp. 163–171. Berlin: Springer.

Azzalini, A. and Bowman, A. W. (1990) A look at some data on the Old Faithful Geyser. *Appl. Statist.*, **39**, 357–365.

Banfield, J. D. and Raftery, A. E. (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics*, **49**, 803–821.

Barbour, A. D. and Reinert, G. (2001) Small worlds. *Rand. Struct. Algs*, **19**, 57–74; correction, **25** (2004), 115.

Barbour, A. D. and Reinert, G. (2006) Discrete small worlds. *Electron. J. Probab.*, **11**, 1234–1283.

Bartholomew, D. J. (1987) *Latent Variable Models and Factor Analysis*. London: Griffin.

Bearman, P., Moody, J. and Stovel, K. (2004) The structure of adolescent romantic and sexual networks. *Am. J. Sociol.*, **110**, 44–91.

Besag, J. (1986) On the statistical analysis of dirty pictures (with discussion). *J. R. Statist. Soc.* B, **48**, 259–302.

Blei, D. M., Ng, A. and Jordan, M. l. (2003) Latent Dirichlet allocation. *J. Mach. Learn. Res.*, **3**, 993–1022.

Boer, P., Huisman, M., Snijders, T. and Zeggelink, E. (2003) StOCNET: an open software system for the advanced statistical analysis of social networks. (Available from `http://stat.gamma.rug.nl/stocnet`.)

Bollobas, B., Janson, S. and Riordan, O. (2006) The phase transition in inhomogeneous random graphs. *Rand. Struct. Algs*, to be published.

Borgatti, S. P., Everett, M. G. and Freeman, L. C. (2002) *UCINET for Windows: Software for Social Network Analysis*. Harvard: Analytic Technologies.

Breiger, R. L. (1974) The duality of persons and groups. *Socl Forces*, **53**, 181–190.

Cartwright, D. and Harary, F. (l956) Structural balance: a generalization of Heider's theory. *Psychol. Rev.*, **63**, 277–292.

Cerioli, A., Riani, M. and Atkinson, A. C. (2006) Robust classification with categorical variables. In *COMPSTAT*

2006: Proc. Computational Statistics (eds A. Rizzi and M. Vichi), pp. 507–519. Heidelberg: Physica.
Davis, J. A. and Leinhardt, S. (1972) The structure of positive interpersonal relations in small groups. In *Sociological Theories in Progress*, vol. 2, pp. 218–251. Boston: Houghton Mifflin.
Doreian, P., Batagelj, V. and Ferligoj, A. (2005) *Generalized Blockmodeling*. Cambridge: Cambridge University Press.
Dorogovtsev, S. N. and Mendes, J. F. F. (2003) *Evolution of Networks: from Biological Nets to the Internet and WWW*. Oxford: Oxford 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.
Durrett, R. (2006) *Random Graph Dynamics*. Cambridge: Cambridge University Press.
Erosheva, E. A. (2003) Bayesian estimation of the grade of membership model. In *Bayesian Statistics 7* (eds J. M. Bernardo, A. P. Dawid, J. O. Berger, M. West, D. Heckerman, M. J. Bayarri and A. F. M. Smith), pp. 501–510. Oxford: Oxford University Press.
Erosheva, E. A., Fienberg, S. E. and Lafferty, J. (2004) Mixed-membership models of scientific publications. *Proc. Natn. Acad. Sci. USA*, **97**, 11885–11892.
Fokoué, E. (2005) Mixtures of factor analyzers: an extension with covariates. *J. Multiv. Anal.*, **95**, 370–384.
Fokoué, E. and Titterington, D. M. (2003) Mixtures of factor analysers: Bayesian estimation and inference by stochastic simulation. *Mach. Learn.*, **50**, 73–94.
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. (2006) `MCLUST` version 3: an R package for normal mixture modeling and model-based clustering. *Technical Report 504*. Department of Statistics, University of Washington, Seattle.
Frank, O. and Strauss, D. (1986) Markov graphs. *J. Am. Statist. Ass.*, **81**, 832–842.
Gelfand, A. E., Kim, H.-J., Sirmans, C. F. and Banerjee, S. (2003) Spatial modeling with spatially varying coefficient processes. *J. Am. Statist. Ass.*, **98**, 387–396.
Gelman, A., Meng, X. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statist. Sin.*, **6**, 753–807.
Gelman, A., Pasarica, C. and Dodhia, R. (2002) Let's practice what we preach: turning tables into graphs. *Am. Statistn*, **56**, 121–130.
Ghahramani, Z. and Beal, M. (2000) Variational inference for Bayesian mixture of factor analysers. In *Advances in Neural Information Processing Systems*, vol. 12 (eds S. A. Solla *et al.*). Cambridge: MIT Press.
Goodreau, S. M. (2006) Birds of a feather, or friend of a friend?: using statistical network analysis to investigate adolescent social networks. *Working Paper*. Center for Studies in Ecology and Demography, University of Washington, Seattle.
Goodreau, S. (2007) Applying advances in exponential random graph (p^*) models to a large social network. *Socl Networks*, to be published.
Gormley, I. C. (2006) Statistical models for rank data. *PhD Thesis*. Trinity College Dublin, Dublin.
Gormley, I. C. and Murphy, T. B. (2005) Exploring irish election data: a mixture modelling approach. *Technical Report 05/08*. Department of Statistics, Trinity College Dublin, Dublin.
Gormley, I. C. and Murphy, T. B. (2006a) A latent space model for rank data. In *Statistical Network Analysis: Models, Issues, and New Directions*. New York: Springer. To be published.
Gormley, I. C. and Murphy, T. B. (2006b) Analysis of Irish third-level college applications data. *J. R. Statist. Soc.* A, **169**, 361–379.
Gormley, I. C. and Murphy, T. B. (2007) Discussion on 'Estimating the integrated likelihood via posterior simulation using the harmonic mean identity' (by A. E. Raftery, M. A. Newton, J. M. Satagopan and P. Krivitsiy). In *Bayesian Statistics 8* (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West). Oxford: Oxford University Press.
Handcock, M. S. (2002) Statistical models for social networks: inference and degeneracy. In *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers* (eds R. Breiger, K. Carley and P. E. Pattison), pp. 229–240. Washington DC: National Academy Press.
Handcock, M. S. (2003) Assessing degeneracy in statistical models of social networks. *Working Paper 39*. Center for Statistics and the Social Sciences, University of Washington, Seattle. (Available from `http://www.csss.washington.edu/Papers/`.)
Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M. and Morris, M. (2003a) statnet: an R package for the statistical modeling of social networks. (Available from `http://www.csde.washington.edu/statnet`.)
Handcock, M. S., Hunter, D. R. and Tantrum, J. (2003b) Latent space models for social network in practice. *Institute for Mathematics and Its Applications Wrkshp 3: Networks and the Population Dynamics of Disease Transmission, Minneapolis*.
Handcock, M. S. and Jones, J. (2004) Likelihood-based inference for stochastic models of sexual network formation. *Theoret. Popln Biol.*, **65**, 413–422.
Handcock, M. S. and Morris, M. (2007) A simple model for complex networks with arbitrary degree distribution and clustering. *Lect. Notes Comput. Sci.*, to be published.
Handcock, M. S., Tantrum J., Shortreed, S. and Hoff, P. (2004) latentnet: an R package for latent position and clus-

ter modeling of statistical networks. (Available from `http://www.csde.washington.edu/statnet/latentnet`.)
Hennig, C. (2004) Breakdown points for maximum likelihood-estimators of location-scale mixtures. *Ann. Statist.*, **32**, 1313–1340.
Hoff, P. D. (2003) Random effects models for network data. In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (eds K. C. R. Breiger and P. Pattison), pp. 303–312. Washington DC: National Academy Press.
Hoff, P. D. (2005) Bilinear mixed-effects models for dyadic data. *J. Am. Statist. Ass.*, **100**, 286–295.
Hoff, P. D. (2007) Multiplicative latent factor models for description and prediction of social networks. *Computnl Math. Organizn Theory*, to be published.
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.
Holland, P. W. and Leinhardt, S. (l970) A method for detecting structure in sociometric data. *Am. J. Sociol.*, **70**, 492–513.
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.
Hunter, D. R., Goodreau, S. M. and Handcock, M. S. (2007) Goodness of fit of social network models. *J. Am. Statist. Ass.*, to be published.
Hunter, D. R. and Handcock, M. S. (2006) Inference in curved exponential family models for networks. *J. Computnl Graph. Statist.*, **15**, 565–583.
Jacobs, R. A., Jordan, M. I., Nowlan, S. J. and Hinton, G. E. (1991) Adaptive mixture of local experts. *Neur. Computn*, **3**, 79–87.
Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *J. Am. Statist. Ass.*, **90**, 773–795.
Kass, R. E. and Wasserman, L. A. (1995) A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J. Am. Statist. Ass.*, **90**, 928–934.
Keribin, C. (2000) Consistent estimation of the order of mixture models. *Sankhya* A, **62**, 49–66.
Kim, H. and Mallick, B. K. (2002) Analyzing spatial data using skew-Gaussian processes. In *Spatial Cluster Modelling* (eds A. B. Lawson and D. Denision), ch 9. London: CRC.
Kossinets, G. and Watts, D. J. (2006) Empirical analysis of an evolving social network. *Science*, **311**, 88–90.
Krackhardt, D. (1988) Predicting with networks: a multiple regression approach to analyzing dyadic data. *Socl Netwrks*, **10**, 359–381.
Lazega, E. and Pattison, P. E. (1999) Multiplexity, generalized exchange and cooperation in organizations: a case study. *Socl Netwrks*, **21**, 67–90.
Lehmann, E. L. (1983) *Theory of Point Estimation*. New York: Wiley.
Leslie, D. S., Kohn, R. and Fiebig, D. G., (2006) Binary choice models with general distributional forms. To be published.
Little, R. J. A. and Rubin, D. B. (1983) On jointly estimating parameters and missing values by maximizing the complete data likelihood. *Am. Statistn*, **37**, 218–220.
Longford, N. T. and Pittau, M. G. (2006) Stability of household income in European countries in the 1990's. *Computnl Statist. Data Anal.*, **51**, 1364–1383.
Luce, R. D. and Perry, A. (1949) A method of matrix analysis of group structure. *Psychometrika*, **51**, 1364–1383.
MacKay, D. J. C. (1995) Bayesian neural networks and density networks. *Instr. Meth. Phys. Res.* A, **354**, 73–80.
Manton, K. G., Woodbury, M. A. and Tolley, D. H. (1994) *Statistical Applications using Fuzzy Sets*. New York: Wiley.
Marriott, F. H. C. (1975) Separating mixtures of normal distributions. *Biometrics*, **31**, 767–769.
Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. and Alon, U. (2004) Superfamilies of evolved and designed networks. *Science*, **303**, 1538–1542.
Murphy, T. B. and Martin, D. (2003) Mixtures of distance-based models for ranking data. *Computnl Statist. Data Anal.*, **41**, 645–655.
Neal, R. M. (2000) Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.*, **9**, 249–265.
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.
Petrone, S. and Raftery, A. E. (1997) A note on the Dirichlet process prior in Bayesian nonparametric inference with partial exchangeability. *Statist. Probab. Lett.*, **36**, 69–83.
Quintana, F. A. and Iglesias, P. L. (2003) Bayesian clustering and product partition models. *J. R. Statist. Soc.* B, **65**, 557–574.

Raftery, A. E., Newton, M. A., Satagopan, J. M. and Krivitsiy, P. (2007) Estimating the integrated likelihood via posterior simulation using the harmonic mean identity (with discussion). In *Bayesian Statistics 8* (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), pp. 1–45. Oxford: Oxford University Press.

Robins, G. L., Pattison, P. E., Kalish, Y. and Lusher, D. (2007a) An introduction to exponential random graph (p*) models for social networks. *Socl Netwrks*, to be published.

Robins, G. L., Snijders, T. A. B., Weng, P., Handcock, M. S. and Pattison, P. E. (2007b) Recent developments in exponential random graph (p*) models for social networks. *Socl Netwrks*, to be published.

Sasik, R., Hwa, T., Iranfar, N. and Loomis, W. F. (2001) Percolation clustering: a novel approach to the clustering of gene expression patterns in Dictyostelium development. In *Proc. Pacific Symp. Biocomputing* (eds R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale and T. E. Klein), pp. 335–347. Singapore: World Scientific Press.

Schweinberger, M. and Snijders, T. A. B. (2003) Settings in social networks: a measurement model. *Sociol. Methodol.*, **33**, 307–341.

Sethuraman, J. (1994) A constructive definition of Dirichlet priors. *Ann. Statist.*, **4**, 639–650.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Shortreed, S., Handcock, M. S. and Hoff, P. (2006) Positional estimation within a latent space model for networks. *Methodology*, **2**, 24–33.

Simmel, G. (1955) *Conflict and the Web of Group Affiliations*. New York: Free Press.

Snijders, T. A. B. (2002) Markov Chain Monte Carlo estimation of exponential random graph models. *J. Socl Struct.*, **3**.

Snijders, T. A. B., Pattison, P. E., Robins, G. L. and Handcock, M. S. (2006) New specifications for exponential random graph models. *Sociol. Methodol.*, **36**, 99–153.

Tobler, W. R. (1976) Spatial interaction patterns. *J. Environ. Syst.*, **6**, 271–301.

Tobler, W. R. (2005) Using asymmetry to estimate potential. *25th Int. Sunbelt Social Network Conf., Redondo Beach*.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*, 4th edn. New York: Springer.

Wang, X. F. and Chen, G. (2003) Complex networks: small-world, scale-free and beyond. *IEEE Circ. Syst. Mag.*, **3**, 6–20.

Wang, Y. J. and Wong, G. Y. (1987) Stochastic blockmodels for directed graphs. *J. Am. Statist. Ass.*, **82**, 8–19.

Watts, D. J., Dodds, P. S. and Newman, M. E. J. (2002) Identity and search in social networks. *Science*, **296**, 1302–1305.

Westveld, A. and Hoff, P. D. (2005) Statistical methodology for longitudinal social network data. In *Proc. A. Meet. American Political Science Association*. American Political Science Association.

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.

Zadeh, L. (1965) Fuzzy sets. *Inform. Control*, **8**, 338–353.

Zijlstra, B. J. H., van Duijn, M. A. J. and Snijders, T. A. B. (2005) Model selection in random effects models for directed graphs using approximated Bayesian factors. *Statist. Neerland.*, **59**, 107–118.