

# Model-Based Clustering for Social Networks

Mark S. Handcock, Adrian E. Raftery and Jeremy M. Tantrum  
University of Washington

Working Paper no. 46  
Center for Statistics and the Social Sciences  
University of Washington

April 27, 2005

## Abstract

Network models are widely used to represent relations among 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* (LPCM), 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 one corresponding to a cluster. We propose two estimation methods: a two-stage maximum likelihood method, and a Bayesian MCMC method; the former is quicker and simpler, but the latter performs better. We also propose a Bayesian way of determining the number of clusters present using approximate conditional Bayes factors. It models 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, 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.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Latent Position Cluster Model (LPCM) for Social Networks</b>	<b>3</b>
<b>3</b>	<b>Estimation</b>	<b>5</b>
3.1	Two-Stage Maximum Likelihood Estimation . . . . .	5
3.2	Bayesian Estimation . . . . .	6
3.3	Identifiability of Positions and Cluster Labels . . . . .	7
3.3.1	Actor Positions via Minimum Kullback-Leibler Divergence . . . . .	8
3.3.2	Label Switching via Minimum Kullback-Leibler Divergence . . . . .	9
<b>4</b>	<b>Choosing the Number of Clusters</b>	<b>10</b>
<b>5</b>	<b>Examples</b>	<b>11</b>
5.1	Example 1: Liking between Monks . . . . .	11
5.2	Example 2: Add Health . . . . .	16
<b>6</b>	<b>Discussion</b>	<b>24</b>

## List of Tables

1	Two-stage maximum likelihood and Bayesian estimates of the parameters of the LPCM for the relationship between monks within a monastery. . . . .	14
2	Add Health data: Clusters from two-stage MLE of the LPCM with 6 clusters, compared to the student's grades. . . . .	19
3	Two-stage MLE and Bayesian estimates of the parameters of the LPCM for the Add Health network. . . . .	20
4	Clusters from the LPCM compared to the student's grades for the Add Health network. Note the concordance between the clusters and the actual grades. . . . .	21

## List of Figures

1	Relationships between monks within a monastery. The groups from two-stage maximum likelihood estimation of the LPCM with three groups are shown by the colors of the nodes. A grouping given by Sampson (1969) is shown by the letters: (T)urks, (L)oyal Opposition, (O)utcasts, and (W)averers. The ties (i.e. the data) are shown by the arrows. . . . .	12
2	BIC plot for the LPCM of the relationships between monks within a monastery . . .	14
3	Estimates of clusters and latent positions for the relationship between monks within a monastery from the Bayesian estimation of the LPCM. The probability of assignment to each latent clusters is shown by a colored pie chart. . . . .	15

4	Two-stage maximum likelihood estimates of the latent positions and clusters for the Add Health data, where the number of clusters (six) is chosen by BIC. Clusters are shown by color with actual grades shown as numbers. There are six green points, representing students from grade 9, two of which are coincident. . . . .	17
5	Clusters from two-stage MLE of the LPCM for the Add Health data, where the number of clusters is constrained to be six. Clusters are shown by color with actual grades shown as numbers. There are six green points, representing students from grade 9, two of which are coincident. . . . .	18
6	BIC plot for the LPCM of the Add Health Network . . . . .	21
7	Bayesian estimates of posterior clusters and latent positions for the friendship network in the Add Health school. Latent clusters are shown by color with actual grades shown as numbers. . . . .	22
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 Add Health school. The students' grades are shown as numbers. . . . .	23

# 1 Introduction

Networks are a useful device to represent relational data, that is, data with properties beyond the attributes of the units (nodes) involved. They can be used to describe the behavior 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 tie or edge represents the presence or absence, or strength of a relationship between the actors. Nodes can be used to represent larger social units (groups, families, organizations), objects (airports, servers, locations), or abstract entities (concepts, texts, tasks, 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  $i, j = 1, \dots, n$ . 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  $Y$ . In the case of binary relations, the data can also be thought of as a graph in which the nodes are actors and the edge set is  $\{(i, j) : y_{i,j} = 1\}$ . When  $(i, j)$  is in the edge set we write  $i \rightarrow j$ .

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, Boorman, and Breiger 1976). Transitivity can lead to some clustering of relationships within the network.

The likelihood of a link 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 and race of the individuals. Homophily by attributes usually implies increased probability of a tie (McPherson, Smith-Lovin, and Cook, 2001), although 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 eyeballing visualizations of the network rather than by more formal inference methods (Liotta 2004).

Existing stochastic models struggle to represent the three common features of social net-

works 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 an edge 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 a 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 and hence clustering (Snijders et al. 2005). Models based only on the degree distributions of the actors, popular in physics and applied mathematics, are also quite restrictive and often do not model clustering at all (Newman 2003).

The seminal work on structural equivalence by Lorrain and White (1971) motivated statistical procedures for clustering or “blocking” relational data (*blockmodels*). This 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, Boorman and Arabie (1975) and White, Boorman, and Breiger (1976) developed and compared alternative algorithms. Subsequent work in this area has been on deterministic algorithms to block actors into prespecified theoretical types (Dorian, Batagelj and Ferligoj 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 blockmodel. 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 attributes.

The idea of representing a social network by assigning positions in a continuous space to

the actors was introduced in the 1970s; see, for example, McFarland and Brown (1973) and Breiger, Boorman and Arabie (1975). These authors 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 has to 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, Raftery and Handcock (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 using standard statistical principles; thus no arbitrary choice of dissimilarity is required. This model automatically represents transitivity and also takes account of homophily on 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 one.

Here we propose a new model, the latent position cluster model (LPCM), that takes account of transitivity, homophily on attributes, and clustering simultaneously in a natural way. It extends the latent space model of Hoff, Raftery and Handcock (2002) to take account of clustering, using the ideas of model-based clustering. 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 MCMC; 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 using approximate conditional Bayes factors. In Section 5 we illustrate the method using two social network datasets.

## 2 The Latent Position Cluster Model (LPCM) for Social Networks

The data we model in this paper 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 analyzed 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:

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

where  $X = \{x_{i,j}\}$  denotes observed characteristics that may be dyad-specific and vector-valued, and  $\beta$  denotes parameters to be estimated.

We model  $P(y_{i,j}|z_i, z_j, x_{i,j}, \beta)$  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:

$$\log \text{odds}(y_{i,j} = 1|z_i, z_j, x_{i,j}, \beta) = \beta'_0 x_{i,j} - \beta_1 |z_i - z_j|. \quad (2)$$

In this way the model accounts for homophily on the attributes  $X$ , as well as transitivity, through the latent space. To identify the scale of the positions and  $\beta_0$  and  $\beta_1$ , we restrict the positions to have unit root mean square:  $\sqrt{\frac{1}{n} \sum_i |z_i|^2} = 1$ .

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

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

where  $\lambda_g$  is the probability that an actor belongs to the  $g$ -th group, so that  $\lambda_g \geq 0$  ( $g = 1, \dots, G$ ) and  $\sum_{g=1}^G \lambda_g = 1$ , 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 model be specified independently of the coordinate system. The model (3) was proposed as a model for clustering 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 maximum likelihood estimation method that first computes the MLE of the (nonclustering) latent space model, and then computes the MLE 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; 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 MLE of the latent positions for the (nonclustering) latent space model of Hoff et al. (2002), as described in that paper. Perhaps surprisingly, this is fairly straightforward because the loglikelihood is convex as a function of the distances between actors, although not as 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 nonlinear optimization method.

The second stage is to find an MLE of the multivariate mixture model conditionally on the latent positions estimated at the first stage. This can be done using the EM algorithm (Dempster, Laird and Rubin 1977). It has been implemented for the model (3) in a clustering context in the R package `mclust` (Fraley and Raftery 1998, 2002, 2003). The likelihood function for the model (3) does not have a unique local maximum, and the one found by the EM algorithm depends 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

We now develop a fully Bayesian approach to estimation of the LPCM given by (1), (2) and (3), using MCMC. We introduce the new variables  $K_i$ , equal to  $g$  if the  $i$ -th actor belongs to the  $g$ -th group, as is standard in Bayesian estimation of mixture models (e.g. Diebolt and Robert 1994).

We use the following prior distributions:

$$\begin{aligned}\beta &\sim N_p(\xi, \Psi), \\ (\lambda_1, \dots, \lambda_G) &\sim \text{Dirichlet}(\nu_1, \dots, \nu_G), \\ \sigma_g^2 &\stackrel{\text{i.i.d.}}{\sim} \sigma_0^2 \text{Inv}\chi_\alpha^2, \quad g = 1 \dots G, \\ \boldsymbol{\mu}_g &\stackrel{\text{i.i.d.}}{\sim} N_d(0, \omega^2 I_d), \quad g = 1 \dots G,\end{aligned}$$

where  $\xi, \Psi, \nu_1, \dots, \nu_G, \sigma_0^2, \alpha$  and  $\omega$  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 allows a wide range of values of  $\beta$ . We take  $\alpha = 2$  and  $\sigma_0^2 = 0.103$  (the 5th percentile of the  $\chi_2^2$  distribution), which implies a prior density on  $\sigma_g^2$  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.

To fit the model using a Bayesian approach, we use a Markov Chain Monte Carlo algorithm, iterating 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 not explicitly specified in the following formulae. The full conditional posterior distributions given the data and these other unknowns are:

$$z_i | K_i = g, \text{ others} \propto \phi_d(z_i; \boldsymbol{\mu}_g, \Sigma_g) P(\mathbf{Y} | Z, \beta), \quad i = 1 \dots n, \quad (4)$$

$$\beta | z_i, \text{ others} \propto \phi_p(\beta; \xi, \Psi) P(Y | Z, \beta), \quad i = 1 \dots n, \quad (5)$$

$$\lambda_1, \dots, \lambda_G | \text{others} \sim \text{Dirichlet}(n_1 + \nu_1, \dots, n_G + \nu_G), \quad (6)$$

$$\boldsymbol{\mu}_g | \text{others} \sim \phi_d \left( \boldsymbol{\mu}_g; \frac{n_g \bar{z}_g}{n_g + 1/\omega^2}, \frac{\sigma_g^2}{n_g + \sigma_g^2/\omega^2} I \right), \quad g = 1 \dots G, \quad (7)$$

$$\sigma_g^2 | \text{others} \sim (\sigma_0^2 + s_g^2) \times \text{Inv}\chi_{\alpha+n_g}^2, \quad g = 1 \dots G, \quad (8)$$

$$P(K_i = g | \text{others}) = \frac{\lambda_g \phi_d(z_i; \boldsymbol{\mu}_g, \Sigma_g)}{\sum_{r=1}^G \lambda_r \phi_d(z_i; \boldsymbol{\mu}_r, \Sigma_r)}, \quad i = 1 \dots n, \quad g = 1 \dots G, \quad (9)$$

where

$$n_g = \sum_{i=1}^n I_{[K_i=g]}, \quad s_g^2 = \frac{1}{d} \sum_{i=1}^n (z_i - \boldsymbol{\mu}_g)^T (z_i - \boldsymbol{\mu}_g) I_{[K_i=g]}, \quad \bar{z}_g = \frac{1}{n_g} \sum_{i=1}^n z_i I_{[K_i=g]},$$

and  $\phi_d(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the  $d$ -dimensional Normal density.

Our algorithm is then as follows:

1. Use Metropolis-Hastings to sample  $Z_{t+1}$ :

(a) Propose  $Z^*$  from the conditional distribution

$$z_i^* | \text{others} \sim \text{MVN}(\mu_{K_i}, \sigma_{K_i}^2).$$

(b) With probability equal to

$$\frac{P(Y|Z^*, \beta_t, X)\pi(Z^*)}{P(Y|Z_t, \beta_t, X)\pi(Z_t)},$$

set  $Z_{t+1} = Z^*$ . Otherwise set  $Z_{t+1} = Z_t$ .

2. Use Metropolis-Hastings to sample  $\beta_{t+1}$ :

(a) Propose  $\beta^*$  from (5).

(b) With probability equal to

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

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

3. Update,  $K_i$ ,  $\mu_g$ ,  $\sigma_g^2$  and  $\lambda_g$  from (6), (7), (8) and (9).

### 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. The latter is called the label-switching problem.

We resolve these nonidentifiabilities (or near nonidentifiabilities in the Bayesian context) by postprocessing 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 so as to resolve the invariance to reflections, rotations and translations, following Oh and Raftery (2001) and Hoff et al. (2002). Second, use the relabelling algorithm of Celeux, Hurn and Robert (2000) to solve the label-switching problem.

Instead, however, we adopt a unified approach which minimizes the Kullback-Leibler divergence between two distributions. The main idea of this is to find the configuration with density closest to the corresponding “true” density. To do this, we postprocess the MCMC sample as follows:

1. Procrustes transform the actor positions, and using the same transformation matrix, transform the cluster means and covariances.
2. Find the minimum Kullback-Leibler positions of the actors.
3. Find the minimum Kullback-Leibler cluster membership probabilities.

We now describe these steps in more detail.

### 3.3.1 Actor Positions via Minimum Kullback-Leibler Divergence

To find the “best” positions of the actors, we minimize the Kullback-Leibler divergence between the density of  $Y$  using the true distances  $\Delta$  and the density of  $Y$  using the position-based distances  $\tilde{\Delta}$ .

Note that  $\Delta_{ij} = \beta^T x_{ij} - |z_i - z_j|$ , with

$$P_{\Delta}(\mathbf{Y} = \mathbf{y}) = \frac{e^{\Delta^T \mathbf{y}}}{c(\Delta)}$$

and the position based distances  $\tilde{\Delta}$  are as described above, with probability  $P_{\tilde{\Delta}}(\mathbf{Y} = \mathbf{y})$ . Then the Kullback-Leibler divergence is

$$\begin{aligned} KL(\Delta, \tilde{\Delta}) &= \sum_{\mathbf{y}} \log \left[ \frac{P_{\Delta}(\mathbf{Y} = \mathbf{y})}{P_{\tilde{\Delta}}(\mathbf{Y} = \mathbf{y})} \right] P_{\Delta}(\mathbf{Y} = \mathbf{y}) \\ &= (\Delta - \tilde{\Delta})^T E_{\Delta}[\mathbf{Y}] - \log \frac{c(\tilde{\Delta})}{c(\Delta)}, \end{aligned}$$

where the sum is over all possible values of  $y$  and

$$E_{\Delta}[\mathbf{Y}]_{ij} = \frac{e^{\Delta_{ij}}}{1 + e^{\Delta_{ij}}}.$$

We choose  $\tilde{\Delta}$  to minimize the posterior mean of  $KL(\Delta, \tilde{\Delta})$ , i.e.

$$E_{\Delta|\mathbf{Y}_{obs}} \left[ E_{\Delta} \left[ KL(\Delta, \tilde{\Delta}) \right] \right] = E_{\Delta|\mathbf{Y}_{obs}} \left[ \Delta E_{\Delta}[\mathbf{Y}] \right] - \tilde{\Delta}^T E[\mathbf{Y}|\mathbf{Y}_{obs}] + E_{\Delta|\mathbf{Y}_{obs}} \left[ \log c(\Delta) \right] - \log c(\tilde{\Delta}),$$

where

$$E[\mathbf{Y}|\mathbf{Y}_{obs}] = E_{\Delta|\mathbf{Y}_{obs}}[E_{\Delta}[\mathbf{Y}]]$$

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

$$\frac{e^{\tilde{\Delta}^T E[\mathbf{Y}|\mathbf{Y}_{obs}]}}{c(\tilde{\Delta})},$$

which can be done using the likelihood maximization method previously described. The posterior mean  $E[\mathbf{Y}|\mathbf{Y}_{obs}]$  can be estimated from the MCMC sample.

### 3.3.2 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  $P_Q(\cdot)$  be the posterior mean of the cluster assignment probability matrix,  $Q$ , where  $q_{ig}$  is the posterior probability of observation  $i$  belonging to cluster  $g$ . Also let  $P_R(\theta)$  be the density of  $R(\theta)$ , where  $r_{ig}(\theta) = Pr(K_i = g|Z, \lambda, \mu, \sigma^2)$ . Minimizing the Kullback-Leibler divergence between  $P_Q(\cdot)$  and  $P_R(\theta)$  with respect to the label switching on the groups will solve the label switching problem.

Let  $v_1, \dots, v_N$  be permutations of the  $N$  samples from the posterior. The permutations of the parameters  $\theta$ , are  $v_t(\theta)$ .

The Kullback-Leibler divergence is

$$KL(R(\theta), Q) = \sum_{i,g} r_{ig}(\theta) \log \frac{r_{ig}(\theta)}{\hat{q}_{ig}},$$

which can be minimized by iterating the following two steps until a fixed point is reached:

Step 1: Choose  $\hat{Q} = (\hat{q}_{ig})$  to minimize

$$\sum_{t=1}^N \sum_{i=1}^n \sum_{g=1}^G r_{ig}(v_t(\theta^{(t)})) \log \frac{r_{ig}(v_t(\theta^{(t)}))}{\hat{q}_{ig}}.$$

Step 2: For  $t = 1, \dots, N$  choose  $v_t$  to minimize

$$\sum_{i=1}^n \sum_{g=1}^G r_{ig}(v_t(\theta^{(t)})) \log \frac{r_{ig}(v_t(\theta^{(t)}))}{\hat{q}_{ig}}.$$

The minimum in Step 1 is achieved by

$$\hat{q}_{ig} = \frac{1}{N} \sum_{t=1}^N r_{ig}(v_t(\theta^{(t)})).$$

## 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.

The simplest approach to Bayesian 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 using the latent space model of Hoff et al. (2002). We then carry out model-based clustering of the resulting estimated latent positions, computing BIC for each different number of groups, and choosing the number of groups with the highest values of BIC, as described by Dasgupta and Raftery (1998) and Fraley and Raftery (2002). As we will see, however, this does not perform too well, and instead we develop an approach 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, obtained by integrating the likelihood times the prior across the model’s parameter space.

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 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 hard to compute the integrated likelihood in a stable way.

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

the integrated likelihood is:

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

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 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 can approximate both of these integrals using the Bayesian Information Criterion (BIC) approximation (Schwartz 1978). The BIC approximation for the integrated likelihood of a model for data  $D$  with  $n_{param}$  parameters  $\theta$  and  $n_{obs}$  observations is:

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

The BIC approximation for logistic regression is:

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

where  $d_{logit}$  is the number of parameters in the logistic model and  $n_{logit}$  is the number of ties, following the ideas of Volinsky and Raftery (2000). The BIC approximation for the mixture model is

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

where  $d_{mbc}$  is the number of parameters in the clustering model and  $n$  is the number of actors. Our final approximation is

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

For both  $\text{BIC}_{lr}$  and  $\text{BIC}_{mbc}$ , we will 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 (Sampson 1969; White, Boorman and Breiger 1976). While resident at the monastery, Sampson

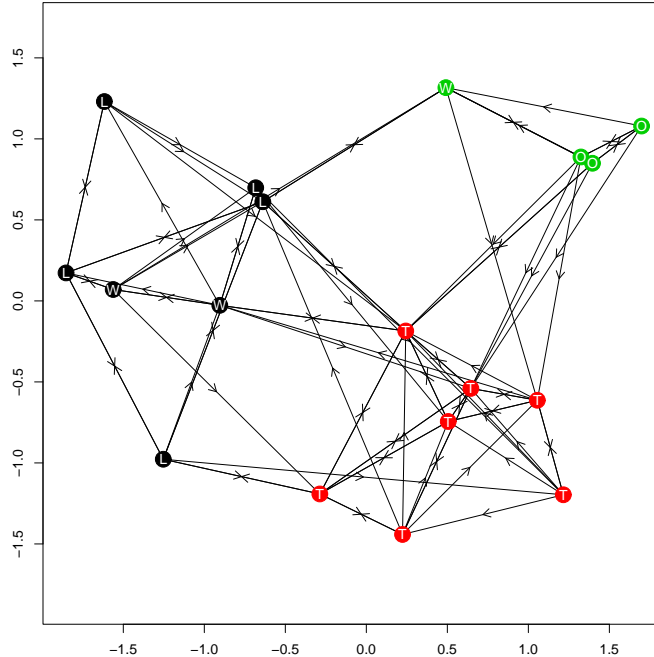


Figure 1: Relationships between monks within a monastery. The groups from two-stage maximum likelihood estimation of the LPCM with three groups are shown by the colors of the nodes. A grouping given by Sampson (1969) is shown by the letters: (T)urks, (L)oyal Opposition, (O)utcasts, and (W)averers. The ties (i.e. the data) are shown by the arrows.

collected extensive sociometric information using interviews, experiments and observation. Here we focus on the social relation of “liking.” We say a monk has the social relation of “like” to another monk if he ranked that monk in the top three monks for positive affection in any of three interviews given over a twelve 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 Figure 1. The BIC criterion from the model-based clustering of the second stage chooses only one cluster. If we ask it to return three clusters we obtain the estimated clusters colored in Figure 1.

The data collected by Sampson (1969) have received a lot of attention in the social networks literature (White, Boorman and Breiger 1976; Wasserman and Faust 1994). Sampson provided a description of the clustering based on information collected at the end of the study period. He identified three main groups: The Young Turks (7 members), the Loyal Opposition (5 members) and the Outcasts (3 members). The other three monks wavered



between the Loyal Opposition and the Young Turks, which he described as being in intense conflict (White, Boorman and Breiger 1976, p.752-753; Sampson 1969, p. 370). The groups identified by Sampson are indicated by letters in Figure 1.

The data we model here include only one of the relationships that Sampson considered in his analysis, and are also from a earlier time period. 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, Boorman and Breiger (1976) developed blockmodels for social relations within the monastery based on eight positive and negative social relations. While 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 three (White, Boorman and Breiger 1976, p. 753). Our model yields the same results as they got, even though they used much more information.

We then fit our Bayesian model using MCMC with 5,000 burn-in iterations that were discarded, and a further 30,000 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 accurate enough 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 Figure 2, and indicates a choice of three clusters. The previous research indicates this to be the correct number of clusters in these data.

Figure 3 shows the minimum Kullback-Leibler estimates of the social positions of the monks for the three cluster model. Here we see that the monks are very 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 clustered with them is the farthest from the others. Overall the Bayesian estimate of the LCPM 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 on Figure 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 Turks.

Table 1: Two-stage maximum likelihood and Bayesian estimates of the parameters of the LPCM for the relationship between monks within a monastery.

	Two-stage	Lower 2.5%	LPCM Posterior Median	Upper 97.5%	Posterior s.d.	Post. median Conditional on $\hat{Z}$
$\beta_0$	3.475	1.028	1.820	2.830	0.458	
$\beta_1$	2.764	1.285	1.756	2.379	0.282	
$\mu_1$	-1.188	-1.753	-1.376	-0.796	0.236	-1.081
$\mu_1$	0.242	-0.407	0.072	0.503	0.235	0.141
$\mu_2$	0.522	-0.170	0.305	0.814	0.250	0.420
$\mu_2$	-0.849	-1.237	-0.772	-0.051	0.301	-0.571
$\mu_3$	1.232	0.292	1.073	1.612	0.335	1.168
$\mu_3$	1.032	0.018	0.686	1.156	0.286	0.756
$\sigma_1$	0.567	0.066	0.217	0.932	0.245	0.233
$\sigma_2$	0.446	0.044	0.137	0.696	0.186	0.152
$\sigma_3$	0.341	0.046	0.696	1.077	0.293	0.156
$\lambda_1$	0.389	0.222	0.389	0.500	0.060	0.389
$\lambda_2$	0.389	0.222	0.389	0.500	0.071	0.389
$\lambda_3$	0.222	0.167	0.222	0.444	0.068	0.222

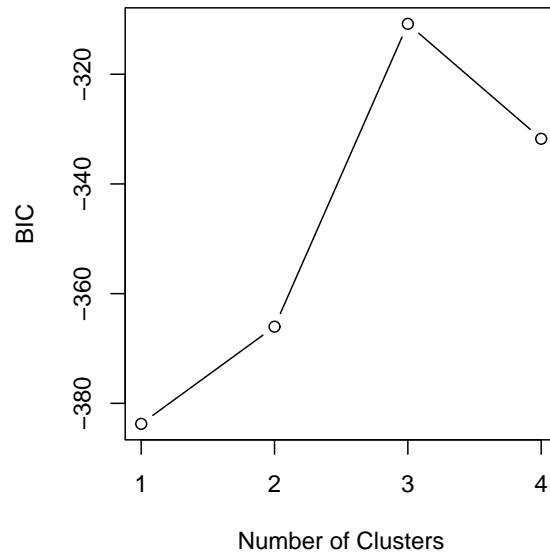


Figure 2: BIC plot for the LPCM of the relationships between monks within a monastery

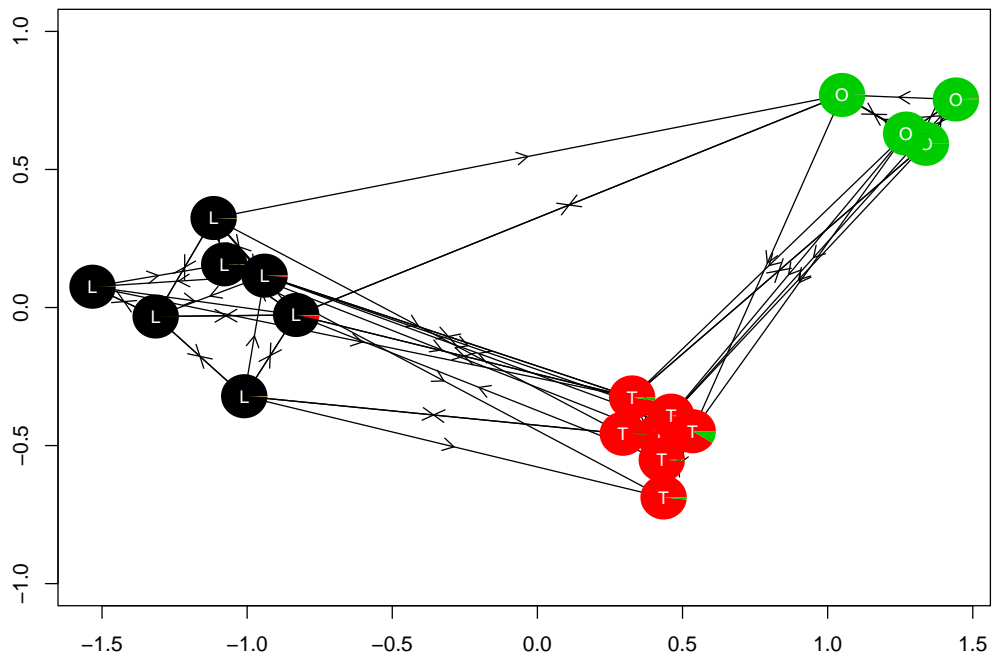


Figure 3: Estimates of clusters and latent positions for the relationship between monks within a monastery from the Bayesian estimation of the LPCM. The probability of assignment to each latent clusters is shown by a colored pie chart.

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: Add Health

The second social network is from the National Longitudinal Study of Adolescent Health, known as the “Add Health” study. Add Health is a school-based, longitudinal study of the health-related behaviors of adolescents and their outcomes in young adulthood. The study design sampled 80 high schools and 52 middle schools from the US representative with respect to region of country, urbanicity, school size, school type, and ethnicity (Harris et al. 2003). In 1994-95 an in-school questionnaire was administered to a nationally representative sample of students in grades 7 through 12. In addition to demographic and contextual information, each respondent was asked to nominate up to five boys and five girls within the school they regarded as their best friends. Thus each student could nominate up to ten students within the school (Udry 2003).

Here we consider a single school of 71 adolescents from grades seven through twelve. 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 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 links.

We fitted the LPCM model without using either the grade or the sex of the adolescents in the model. We will use these instead for assessing the clustering. The two-stage MLEs of the latent positions are given in Figure 4. The approximate BIC values based on the two-stage MLEs chose two clusters, which seems a poor choice given the grade and sex information. When we required six clusters, we got the results shown in Figure 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 7th and 8th grade adolescents belong to two clusters that are mostly homogeneous with respect to grade. The 9th grade adolescents fall into two clusters. The 10th, 11th and 12th grade adolescents fall into two clusters, one of which seems to have all four 12th graders.

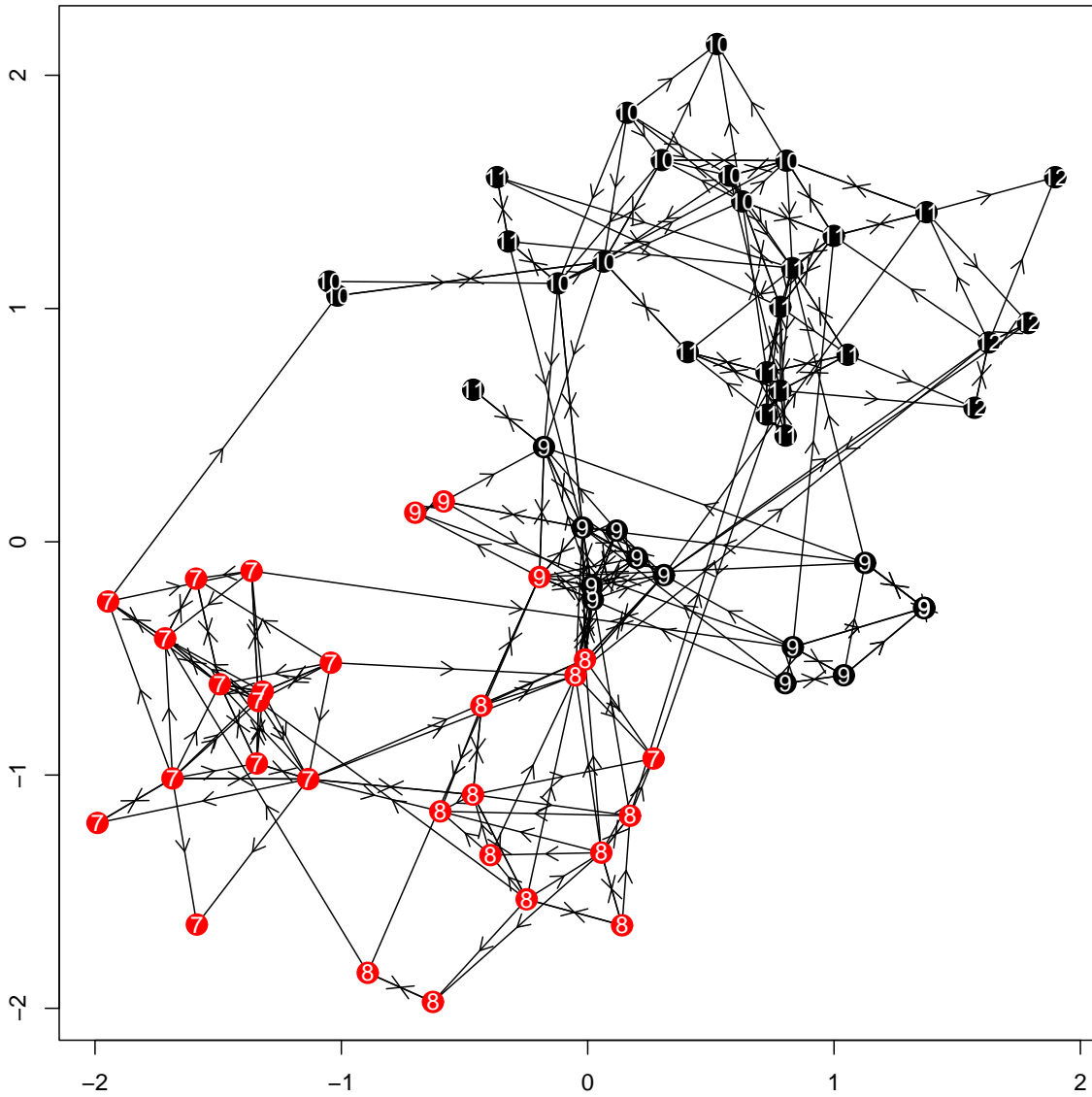


Figure 4: Two-stage maximum likelihood estimates of the latent positions and clusters for the Add Health data, where the number of clusters (six) is chosen by BIC. Clusters are shown by color with actual grades shown as numbers. There are six green points, representing students from grade 9, two of which are coincident.

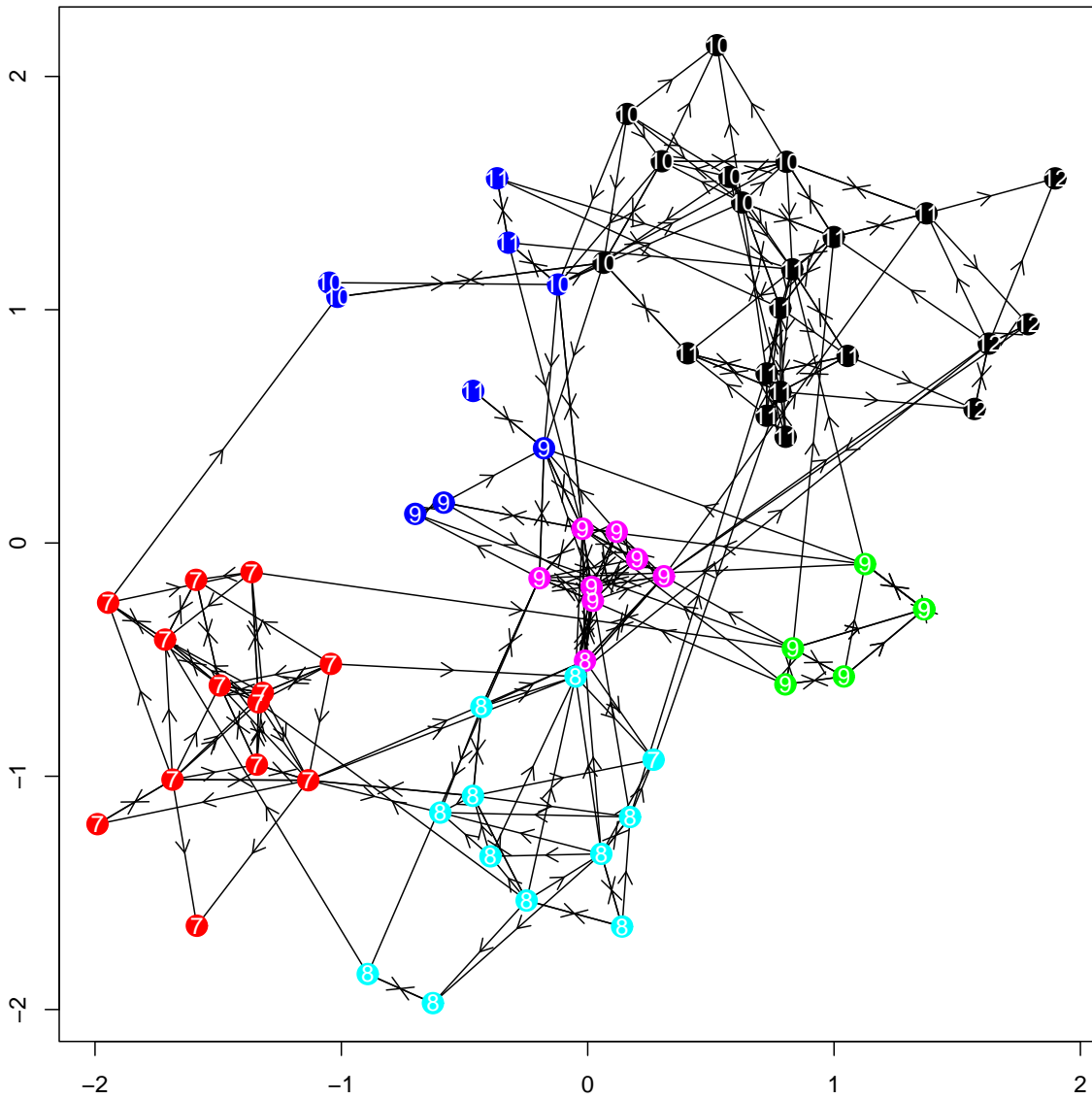


Figure 5: Clusters from two-stage MLE of the LPCM for the Add Health data, where the number of clusters is constrained to be six. Clusters are shown by color with actual grades shown as numbers. There are six green points, representing students from grade 9, two of which are coincident.

Table 2: Add Health data: Clusters from two-stage MLE of the LPCM with 6 clusters, compared to the student’s grades.

grade	Cluster					
	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

We fit our Bayesian model using MCMC with 50,000 burn-in iterations that were discarded, and a further two million iterations, of which we kept every 1000th value. The resulting BIC (Figure 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 Bayesian estimates of the LPCM are shown in Table 3. Figure 7 shows the Bayesian estimates of the children’s social positions for the six-cluster model. As the grades of the students are expected to be an important determinant of social structure, we can compare to the fit of the model to that social structure determined by the grade. 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 7th grade adolescents are in their own well separated cluster, with the exception of one 7th grader whose only friends are 8th graders (possibly a student who had been held back). The 8th graders are mostly in their own cluster, with two having stronger ties to the 9th grade class and so being incorporated into that cluster. The 9th grade class is split into two clusters, the social magentas and the cliquy greens. The magenta 9th grade cluster has many ties to other clusters, whereas the ties from the green 9th grade cluster to other clusters are mostly to the other (magenta) 9th grade cluster. The 11th grader whose only friend is a green 9th grader is more likely to belong to the green cluster 9th grade than to any other. 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 of the 12th graders are in the black cluster. Thus the clustering model has identified within-grade heterogeneity and between-grade homogeneity. This indicates that grade is a primary determinant of social structure but additional social

Table 3: Two-stage MLE and Bayesian estimates of the parameters of the LPCM for the Add Health network.

	LP-MBC	Lower 2.5%	LPCM Posterior Median	Upper 97.5%	Posterior s.d.	Conditional on $\hat{Z}$
$\beta_0$	1.031	1.001	1.394	1.812	0.208	
$\beta_1$	5.205	3.464	3.962	4.549	0.276	
$\mu_{11}$	-1.495	-1.579	-1.168	-0.485	0.271	-1.126
$\mu_{12}$	-0.704	-1.493	-0.532	0.091	0.405	-0.610
$\mu_{21}$	-0.278	-1.130	-0.348	0.385	0.378	-0.241
$\mu_{22}$	-1.276	-1.574	-1.048	0.177	0.448	-1.163
$\mu_{31}$	0.051	-0.484	0.196	0.758	0.304	0.180
$\mu_{32}$	-0.165	-0.819	-0.179	0.526	0.315	-0.224
$\mu_{41}$	1.088	0.008	0.999	1.546	0.391	1.028
$\mu_{42}$	-0.384	-0.940	-0.276	0.594	0.385	-0.392
$\mu_{51}$	-0.497	-0.430	0.349	1.019	0.364	0.599
$\mu_{52}$	0.808	-0.456	1.243	1.683	0.317	1.118
$\mu_{61}$	0.837	-0.863	-0.057	1.052	0.490	-0.172
$\mu_{62}$	1.139	-0.310	0.830	1.476	0.464	0.880
$\sigma_1$	0.368	0.169	0.368	0.895	0.216	0.231
$\sigma_2$	0.402	0.192	0.495	1.012	0.267	0.257
$\sigma_3$	0.166	0.123	0.291	1.035	0.325	0.107
$\sigma_4$	0.204	0.178	0.456	1.022	0.294	0.207
$\sigma_5$	0.450	0.262	0.538	0.985	0.448	0.257
$\sigma_6$	0.514	0.152	0.479	1.064	0.323	0.233
$\lambda_1$	0.188	0.058	0.159	0.348	0.070	0.188
$\lambda_2$	0.174	0.043	0.159	0.348	0.075	0.159
$\lambda_3$	0.116	0.029	0.116	0.333	0.072	0.130
$\lambda_4$	0.087	0.043	0.130	0.319	0.070	0.145
$\lambda_5$	0.130	0.058	0.217	0.406	0.094	0.188
$\lambda_6$	0.304	0.029	0.145	0.348	0.085	0.188



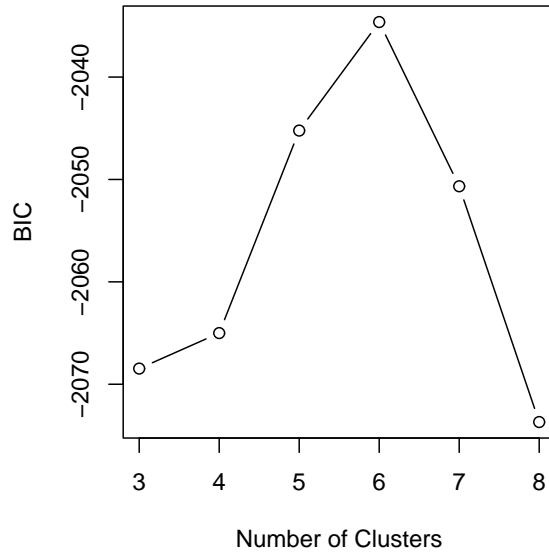


Figure 6: BIC plot for the LPCM of the Add Health Network

forces lead to cohesion and separation beyond that captured by grade.

Figure 8 shows the cluster assignment probabilities for each student. Students in the magenta cluster also have a significant probability of belonging to the cyan 8th grade cluster, whereas the magenta 9th graders only have have significant probability of belonging to the magenta and green clusters. The uncertainty in cluster assignment between the blue and black clusters is clearly visible.

The Bayesian model provides a much better estimate of the number of groups than the two-stage MLE approach. The clusters are well defined in terms of both their positions

Table 4: Clusters from the LPCM compared to the student’s grades for the Add Health network. Note the concordance between the clusters and the actual grades.

grade	Cluster					
	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

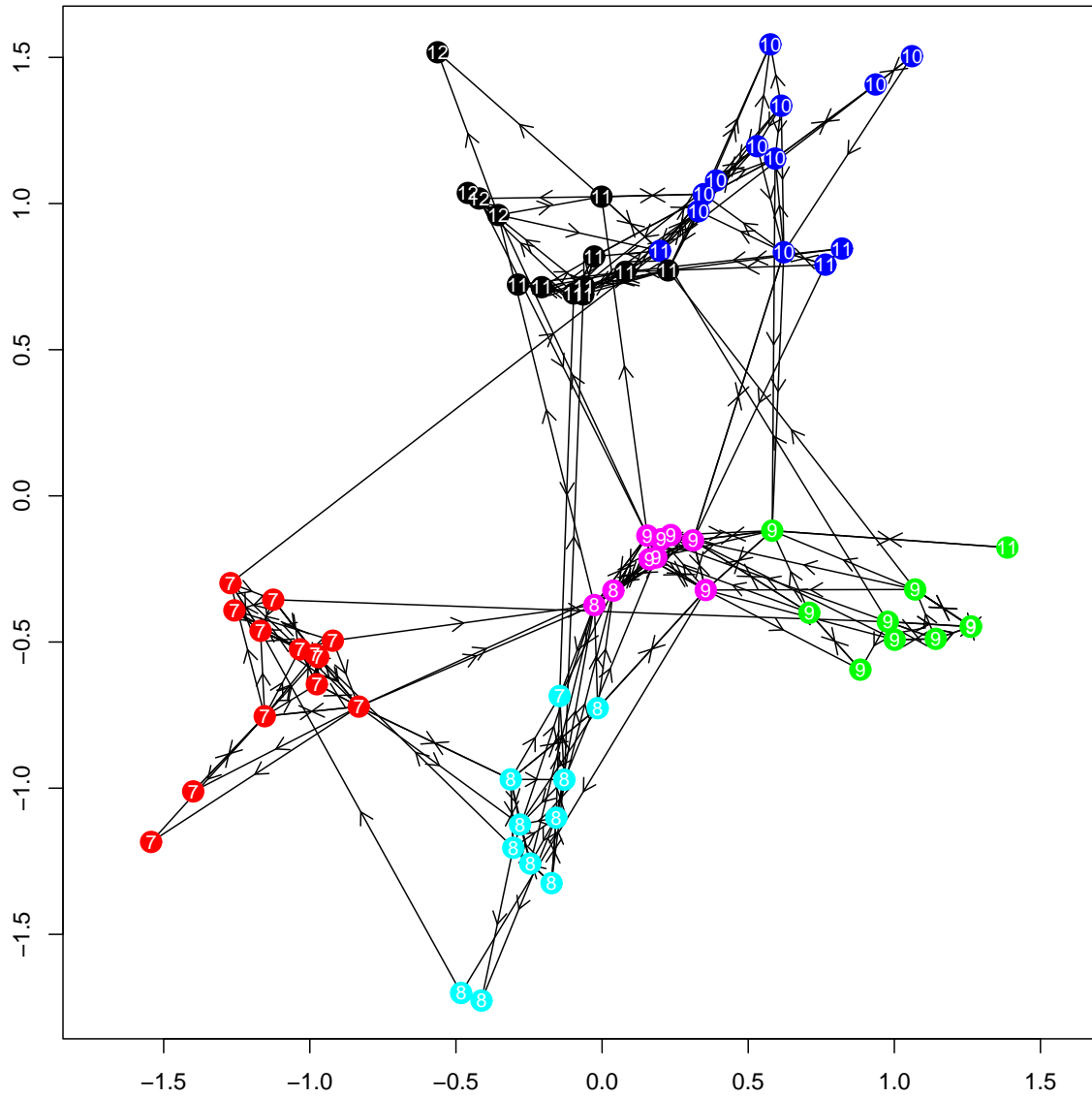


Figure 7: Bayesian estimates of posterior clusters and latent positions for the friendship network in the Add Health school. Latent clusters are shown by color with actual grades shown as numbers.

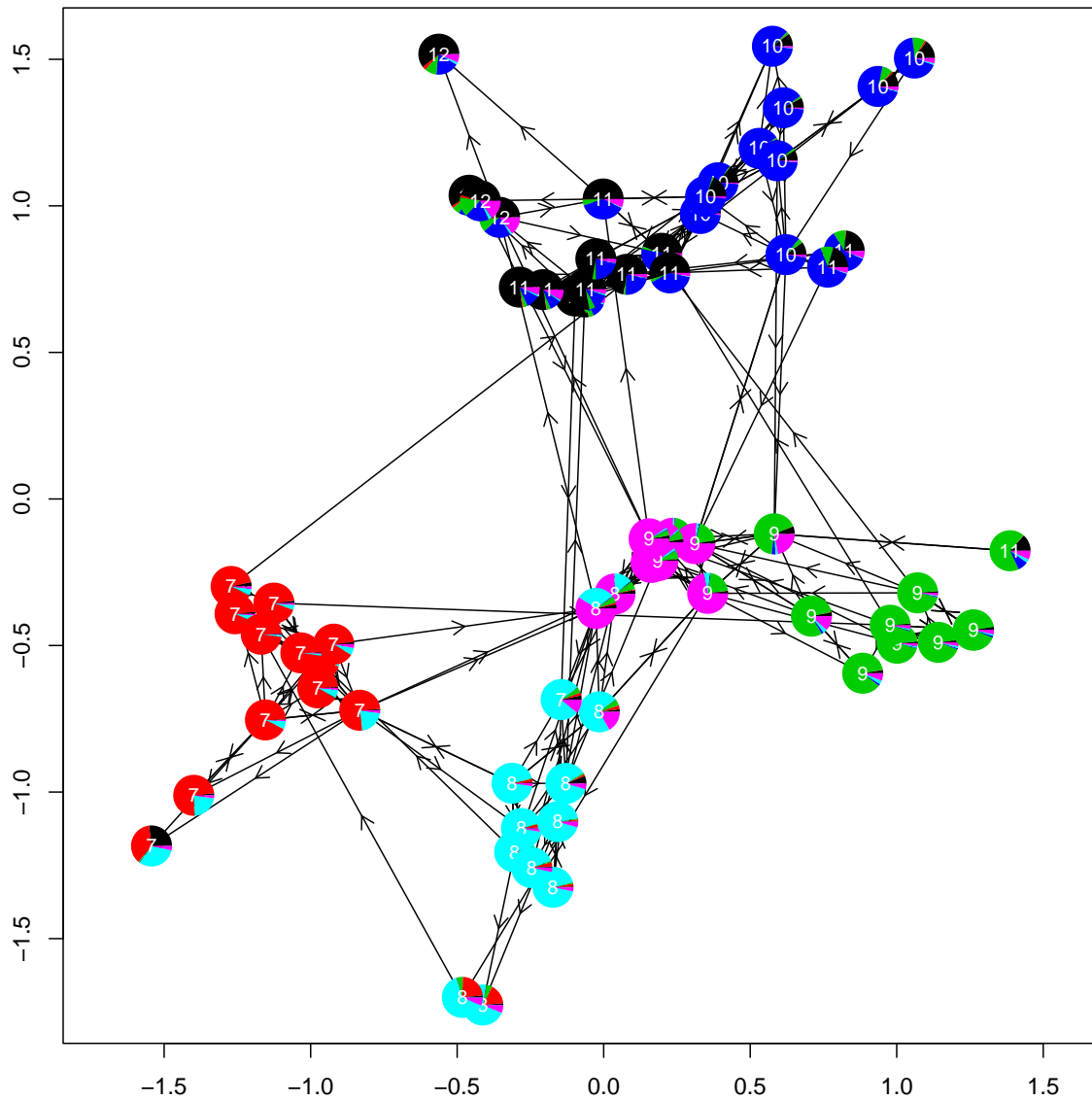


Figure 8: Pie charts for posterior probabilities of cluster assignment for each actor, at the Bayesian estimates of posterior latent positions for the friendship network in the Add Health school. The students' grades are shown as numbers.

in space and their correspondence to the grades. This is reflected in the estimates of the positions and the uncertainty of cluster membership (Figure 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: 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. 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 MLE approach works fairly well and is simple to implement, while the fully Bayesian approach performs better, but is more complex.

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 is only one, for the unordered pair  $(i, j)$ . We have specified our model as one for binary ties: present or absent. However, network ties often have nonbinary 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 (2) by a generalized linear model or another specification of dependence.

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 (2), similarly to 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 well enough, 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 modeling of the spread of contagious diseases (Kretzschmar and Morris 1996; Bearman, Moody, and Stovel 2004; Eames and Keeling 2004; Eubank, Guclu, Kumar, Marathe, Srinivasan, Toroczkai, and Wang 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 analyzed. While feasible, our method is computationally demanding, and so for larger networks more computationally efficient versions of our estimation methods should be sought.

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 using methods similar 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 adequately represent the network, especially for larger networks.

An R package implementing the procedures in this paper is available from the authors, and will be made publicly available on CRAN.

## Acknowledgements

The research of Tantrum and Handcock was supported by NIDA Grant DA012831 and NICHD Grant HD041877. Raftery’s research was supported by NIH grant 8 R01EB 002137–02.

## References

- Anderson, C. J. and S. S. Wasserman (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Social Networks* 9, 1–36.
- Banfield, J. D. and A. E. Raftery (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics* 49, 803–821.
- Bearman, P. S., J. Moody, and K. Stovel (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology* 110, 44–91.
- Breiger, R. L., S. A. Boorman, and P. Arabie (1975). An algorithm for clustering relational data with application to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology* 12, 328–383.
- Celeux, G., M. Hurn, and C. Robert (2000). Computational and inferential difficulties with mixture posterior distribution. *Journal of the American Statistical Association* 95, 957–970.

- Dasgupta, A. and A. E. Raftery (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association* *93*, 294–302.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* *39*, 1–38.
- Diebolt, J. and C. Robert (1994). Bayesian estimation of finite mixture distributions. *Journal of the Royal Statistical Society, Series B* *56*, 363–375.
- Doreian, P., V. Batagelj, and A. Ferligoj (2005). *Generalized Blockmodeling*. Cambridge: Cambridge University Press.
- Eames, K. T. D. and M. J. Keeling (2004). Monogamous networks and the spread of sexually transmitted diseases. *Mathematical Biosciences* *189*, 115–130.
- Eubank, S., H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang (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. *Social Networks* *10*, 313–341.
- Fienberg, S. E. and S. S. Wasserman (1981). Categorical data analysis of single sociometric relations. *Sociological Methodology* *11*, 156–192.
- Fraley, C. and A. E. Raftery (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal* *41*, 578–588.
- Fraley, C. and A. E. Raftery (2002). Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association* *97*, 611–631.
- Fraley, C. and A. E. Raftery (2003). Enhanced model-based clustering, density estimation and discriminant analysis software: MCLUST. *Journal of Classification* *20*, 263–286.
- Frank, O. and D. Strauss (1986). Markov graphs. *Journal of the American Statistical Association* *81*, 832–842.
- Harris, K. M., F. Florey, J. Tabor, P. S. Bearman, J. Jones, and R. J. Udry (2003). The national longitudinal of adolescent health: Research design [www document]. Technical report, Carolina Population Center, University of North Carolina at Chapel Hill, Available at: <http://www.cpc.unc.edu/projects/addhealth/design>.

- Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association* 100, 286–295.
- Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97, 1090–1098.
- Hoff, P. D. and M. D. Ward (2004). Modeling dependencies in international relations networks. *Political Analysis* 12, 160–175.
- Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association* 76, 33–65.
- Kass, R. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773–795.
- Kretzschmar, M. and M. Morris (1996). Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences* 133, 165–195.
- Liotta, G. (Ed.) (2004). *Graph Drawing, 11th International Symposium, GD 2003, Perugia, Italy, September 21-24, 2003, Revised Papers*, Volume 2912 of *Lecture Notes in Computer Science*. Springer.
- Lorrain, F. and H. White (1971). Structural equivalence of individuals in social networks blockstructures with covariates. *Journal of Mathematical Sociology* 1, 49–80.
- McFarland, D. D. and D. J. Brown (1973). Social distance as a metric: a systematic introduction to smallest space analysis. In E. O. Laumann (Ed.), *Bonds of Pluralism: The Form and Substance of Urban Social Networks*, pp. 213–253. New York: Wiley.
- McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *American Review of Sociology* 27, 415–444.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review* 45, 167–256.
- Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96, 1077–1087.
- Oh, M. S. and A. E. Raftery (2001). Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association* 96, 1031–1044.
- Oh, M. S. and A. E. Raftery (2003). Model-based clustering with dissimilarities: A Bayesian approach. Technical Report 441, Department of Statistics, University of

Washington.

- Raftery, A. E. and S. M. Lewis (1996). Implementing MCMC. In W. R. Gilks, D. J. Spiegelhalter, and S. Richardson (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 115–130. London: Chapman and Hall.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 497–511.
- Schweinberger, M. and T. A. B. Snijders (2003). Settings in social networks: A measurement model. *Sociological Methodology* 33, 307–341.
- Sibson, R. (1979). Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the Royal Statistical Society, Series B* 41, 217–229.
- Snijders, T. A. B. and K. Nowicki (1997). Estimation and prediction for stochastic blockstructures for graphs with latent block structure. *Journal of Classification* 14, 75–100.
- Snijders, T. A. B., P. Pattison, G. L. Robins, and M. S. Handcock (2004). New specifications for exponential random graph models. Working Paper 42, Center for Statistics and the Social Sciences, University of Washington.
- Tallberg, C. (2005). A Bayesian approach to modeling stochastic blockstructures with covariates. *Journal of Mathematical Sociology* 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 [machine-readable data file and documentation]. Technical report, Carolina Population Center, University of North Carolina at Chapel Hill.
- Volinsky, C. T. and A. E. Raftery (2000). Bayesian information criterion for censored survival models. *Biometrics* 56, 256–262.
- Wasserman, S. S. and K. Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- White, H. C., S. A. Boorman, and R. L. Breiger (1976). Social-structure from multiple networks I: Blockmodels of roles and positions. *American Journal of Sociology* 81, 730–780.