# Representing Degree Distributions, Clustering, and Homophily in Social Networks With Latent Cluster Random Effects Models

Pavel N. Krivitsky Mark S. Handcock Adrian E. Raftery
Peter D. Hoff[1]
University of Washington

Technical Report no. 517
Department of Statistics
University of Washington

August 7, 2007

[1]Pavel N. Krivitsky is Graduate Student, Mark S. Handcock is Professor and Chair, Adrian E. Raftery is Blumstein-Jordan Professor of Statistics and Sociology, and Peter D. Hoff is Associate Professor, all at the Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322. The research of Krivitsky and Handcock was supported by NIDA Grant DA012831 and NICHD Grant HD041877. The research of Krivitsky and Raftery was supported by NIH grant 8 R01EB 002137–02. Raftery thanks Miroslav Kárný and the Institute of Information Theory and Automation, Prague, as well as Gilles Celeux and INRIA, France, for hospitality during the preparation of this paper.

**Abstract**

Social network data often involve transitivity, homophily on observed attributes, clustering, and heterogeneity of actors. We propose the latent cluster random effects model to take account of all of these features, and we describe a Bayesian estimation method. The model fits two real datasets well. We show by simulation that networks with the same degree distribution can have very different clustering behaviors. This suggests that scale-free and small-world network models may not be adequate for all types of network, while our model recovers both the clustering and the degree distribution.

# 1 Introduction

Social network data consist of data about pairs of actors or nodes. Often these data represent the presence, absence or value of a relationship between pairs of actors, such as liking, respect, familial relationship, shared membership in a group of individuals, or volume of trade for collectivities such as countries or companies. Here we consider binary social network data, representing presence or absence of a relationship.

Much social network data share a number of features. One of these is *transitivity*, for example the fact that if actor $A$ relates to actor $B$ and actor $B$ relates to actor $C$, then actor $A$ is more likely to relate to actor $C$. Another is *homophily on observed attributes*, according to which actors with similar characteristics are more likely to relate. A third feature is *clustering*, according to which actors cluster into unobserved groups, within which links are more likely. This can be due to social self-organization or to homophily on unobserved attributes (such as, for example, interest in the same sport, about which the analyst might not have information). A fourth feature is *heterogeneity*, namely the tendency of some actors to send and/or receive links more than others.

Hoff, Raftery, and Handcock (2002) proposed the latent space model for social networks. This postulates an unobserved Euclidean social space in which each actor has a position. The probability of a link between pairs of actors depends on the distance between them in the space and on their observed characteristics. Estimation of the model involves estimating both the latent positions and the parameters of the model specifying how the probability of a link depends on distance and observed attributes. This accounts for transitivity automatically through the latent space and is flexible enough to include the other common features of social network data also. This was extended by Handcock, Raftery, and Tantrum (2007) — hereafter HRT — to include model-based clustering of the latent space positions, giving a way to detect groups of actors. Separately, Hoff (2005) added random sender and receiver effects to model inhomogeneity of the actors, similarly to those in the $p_2$ model (van Duijn, Snijders, and Zijlstra, 2004).

No model so far proposed has modeled all the four common features of social network data that we mentioned above. In this paper we propose the Latent Cluster Random Effects Model, which explicitly models all four features by adding the random sender and receiver effects as proposed by Hoff (2005) to the latent position cluster model of HRT.

Heterogeneity of actors in social networks has often been modeled via the degree distribution and assuming the network is scale free. These models assume that all networks with the same degree distribution are equally likely. We show through a number of simulated examples that the latent cluster random effects model can model such heterogeneity effectively. We also show that networks with the same degree distribution can have very different clustering behavior, suggesting that scale-free and small-world networks are not adequate to model all networks with the same degree distribution.

In Section 2 we introduce the latent cluster random effects model and in Section 3 we describe our Bayesian method for estimating it using Markov chain Monte Carlo. In Section 4 we illustrate the method using two social network datasets and two simulated datasets with the same degree distribution but different clustering behavior.

# 2 The Latent Cluster Random Effects Model for Social Networks

We first review the latent position cluster model of HRT and then expand it to allow for actor-specific random effects. The data we model consist of $y_{i,j}$, the value of the relation from actor $i$ to actor $j$ for each dyad consisting of two of the $n$ actors. These form the elements of the $n \times n$ sociomatrix $Y$. There may also be dyadic-level covariate information represented by $p$ matrices $X = \{X_k\}_{k=1}^{p} \in \mathbb{R}^{n\times n\times p}$. 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.

The model posits that each actor $i$ has an unobserved position, $Z_i$, in a $d$-dimensional Euclidean latent social space, as in Hoff et al. (2002) and HRT. We then assume that the tie values are stochastically independent given the distances between the actors' positions. Specifically:

$$\text{logit}\left(p(Y_{i,j} = 1|Z, X, \beta)\right) = \eta_{i,j} = \sum_{k=1}^{p} \beta_k X_{k,i,j} - ||Z_i - Z_j||, \tag{1}$$

where $\text{logit}(p) = \log(p/(1-p))$ and $\beta$ denotes the parameters to be estimated. The model accounts for transitivity, through the latent space, as well as homophily on the observed attributes $X$. As in HRT, we allow for any clustering in the $Z_i$ via a finite spherical multivariate normal mixture:

$$Z_i \overset{\text{i.}i.d.}{\sim} \sum_{g=1}^{G} \lambda_g \text{MVN}_d(\mu_g, \sigma_g^2 I_d) \quad i = 1, \ldots, n,$$

where $\lambda_g$ is the probability that an actor belongs to the $g$-th group, so that $\lambda_g \geq 0$ ($g = 1, \ldots, G$) and $\sum_{g=1}^{G} \lambda_g = 1$, and $I_d$ is the $d \times d$ identity matrix. Thus the positions of the actors are drawn from one of $G$ groups, each one centered on a different mean and dispersed with a different variance. The simplification provided by the spherical covariance matrices is motivated by the symmetry of the latent social space, although more structure can be added (see the discussion of HRT).

To represent heterogeneity in the propensity for actors to form ties not captured by the dyad-level covariates or actor positions, we introduce actor-specific random effects. The nature of the effects differ for directed and undirected relationships. For an undirected relationship, each actor $i$ has a latent "sociality" denoted by $\delta_i$, representing his or her propensity to form ties with other actors. The effect of these random effects on the propensity to form ties is modeled as:

$$\eta_{i,j} = \sum_{k=1}^{p} \beta_k X_{k,i,j} - ||Z_i - Z_j|| + \delta_i + \delta_j. \tag{2}$$

The sociality $\delta_i$ is then the conditional log-odds ratio of an actor $i$ having a tie with another actor compared to an actor with similar position and covariates but having $\delta = 0$.

This model can also be used for directed relationships. In this case we define both sender and receiver random effects, $\delta_i$ and $\gamma_i$, representing actor $i$'s propensity to send and receive links, respectively. The model then becomes:

$$\eta_{i,j} = \sum_{k=1}^{p} \beta_k X_{k,i,j} - ||Z_i - Z_j|| + \delta_i + \gamma_j, \tag{3}$$

where

$$\begin{aligned} \delta_i &\overset{\text{i.}i.d.}{\sim} \mathrm{N}(0, \sigma_\delta^2) \quad i = 1, \ldots, n, \\ \gamma_i &\overset{\text{i.}i.d.}{\sim} \mathrm{N}(0, \sigma_\gamma^2) \quad i = 1, \ldots, n, \end{aligned}$$

and the variances $\sigma_\delta^2$ and $\sigma_\gamma^2$ measure the heterogeneity in propensity to send and receive links. The use of random effects in the latent space model was proposed by Hoff (2003), and van Duijn et al. (2004) who made a similar proposal for the $p_2$ model.

# 3 Estimation

## 3.1 Bayesian Estimation and Prior Distributions

We propose a Bayesian approach to estimate the latent cluster random effects model given by (1) and (2), or by (3). The approach estimates the latent positions, the clustering model and the actor-specific effects simultaneously. We implement the methods computationally using a Markov chain Monte Carlo (MCMC) algorithm.

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 (Diebolt and Robert, 1994).

We specify prior distributions as follows:

$$
\begin{aligned}
\beta &\sim \text{MVN}_p(\xi, \Psi), \\
\lambda &\sim \text{Dirichlet}(\nu), \\
\sigma^2_\delta &\sim \alpha_\delta \sigma^2_{0,\delta} \text{Inv}\chi^2_{\alpha_\delta}, \\
\sigma^2_\gamma &\sim \alpha_\gamma \sigma^2_{0,\gamma} \text{Inv}\chi^2_{\alpha_\gamma}, \\
\sigma^2_g &\overset{\text{i.}i.d.}{\sim} \alpha_Z \sigma^2_{0,Z} \text{Inv}\chi^2_{\alpha_Z} \quad g = 1, \ldots, G, \\
\mu_g &\overset{\text{i.}i.d.}{\sim} \text{MVN}_d\left(0, \omega^2 I_d\right), \quad g = 1 \ldots G,
\end{aligned}
$$

where $\xi$, $\Psi$, $\nu = (\nu_1, \ldots, \nu_G)$, $\sigma^2_{0,Z}$, $\alpha_Z$, $\sigma^2_{0,\delta}$, $\alpha_\delta$, $\sigma^2_{0,\gamma}$, $\alpha_\gamma$, and $\omega^2$ are hyperparameters to be specified by the user.

We set $\nu_g$ to the smallest group size we are willing to consider for the network of interest, and $\xi = 0$ and $\Psi = 9I$, which allows a wide range of values of $\beta$. The other hyperparameters are not so clear-cut. In general, larger networks with larger groups and more structure tend to have greater prior within-cluster and between-cluster variances.

In particular, we used $\sigma^2_{0,Z} = \frac{1}{8} \sqrt[d/2]{\frac{n}{G}}$, $\alpha_Z = \sqrt{\frac{n}{G}}$, $\omega^2 = \frac{1}{4} \sqrt[d/2]{n}$, and $\nu_g = \sqrt{\frac{n}{G}}$. Heuristically, networks with larger clusters call for greater prior variances, and it is helpful to have slightly stronger priors for larger clusters, but as a network gets larger, the role of the prior variances in determining the posterior variances should decline. The hyperparameter choices used reflect these intuitions.

## 3.2 Markov chain Monte Carlo algorithm

Our MCMC algorithm iterates over the model parameters with the priors given above, the latent positions $z_i$, the random effects $\delta_i$ and $\gamma_i$, and the group memberships $K_i$. We update variables in turn, and block-update those we expect to be highly correlated. For those variables for which a conjugate prior was specified, full conditional updates are used. The others are updated using Metropolis-Hastings. We describe these in turn.

We first describe the full conditional updates. Let ellipsis ("...") represent those variables from which the variable being sampled is separated, and which thus do not figure in its full conditional distribution. The relevant priors being conjugate, the full conditionals for those variables that can be Gibbs-sampled are as follows:

$$
\begin{aligned}
\sigma^2_\delta|\delta,\ldots &\sim \left(\alpha_\delta\sigma^2_{0,\delta}+\sum_{i=1}^n\delta_i^2\right)\mathrm{Inv}\chi^2_{\alpha_\delta+n},\\
\sigma^2_\gamma|\gamma,\ldots &\sim \left(\alpha_\gamma\sigma^2_{0,\gamma}+\sum_{i=1}^n\gamma_i^2\right)\mathrm{Inv}\chi^2_{\alpha_\gamma+n},\\
\mu_g|Z,K,\sigma_g^2,\ldots &\overset{\text{ind}}{\sim} \mathrm{MVN}_d\left(\frac{n_g\bar{Z}_g}{n_g+\sigma_g^2/\omega^2},\frac{\sigma_g^2}{n_g+\sigma_g^2/\omega^2}\right)\quad g=1,\ldots,G,\\
\sigma_g^2|Z,K,\mu_g,\ldots &\overset{\text{ind}}{\sim} \left(\alpha_Z\sigma^2_{Z,0}+SS_{Z_g}\right)\mathrm{Inv}\chi^2_{\alpha_Z+n_gd}\quad g=1,\ldots,G,\\
\lambda|K,\ldots &\sim \mathrm{Dirichlet}\left(\nu_1+n_1,\ldots,\nu_G+n_G\right),\\
\Pr\left(K_i=g|\lambda,Z,\mu_g,\sigma_g^2,\ldots\right) &\overset{\text{ind}}{=} \frac{\lambda_g f_{\mathrm{MVN}_d\left(\mu_g,\sigma_g^2I_d\right)}(Z_i)}{\sum_{k=1}^G\lambda_k f_{\mathrm{MVN}_d\left(\mu_k,\sigma_k^2I_d\right)}(Z_i)}\quad i=1,\ldots,n,
\end{aligned}
$$

where $SS_{Z_g}=\sum_{i=1}^n 1_{K_i=g}\left(Z_i-\mu_g\right)^T\left(Z_i-\mu_g\right)$, the sum of squared deviations of the latent positions in cluster $g$ from their cluster's mean, and $n_g=\sum_{i=1}^n 1_{K_i=g}$, the number of actors assigned to cluster $g$ during a particular iteration.

We now describe the Metropolis-Hastings updates. Two kinds of Metropolis-Hastings proposals are used. First, actor-specific parameters (latent space positions and random effects) are updated one actor at a time, in a random order. Second, covariate coefficients are block-updated with the scale of latent space positions and a shift in random effects.

An independent $d$-variate normal jump is proposed for each actor (in random order). For a particular actor $i$, the proposal

$$Z_i^*\sim\mathrm{MVN}_d\left(Z_i,\tau_Z^2I_d\right)$$

is made. At the same time, an independent proposal is made for the sender and receiver effects of that actor:

$$
\begin{aligned}
\delta_i^* &\sim \mathrm{N}\left(\delta_i,\tau_\delta^2\right),\\
\gamma_i^* &\sim \mathrm{N}\left(\gamma_i,\tau_\gamma^2\right).
\end{aligned}
$$

The parameters $Z_i^*$, $\delta_i^*$, and $\gamma_i^*$ are then accepted or rejected as a block. The reason for this block-updating is that parameters pertaining to a particular node are likely to be associated. This proposal is symmetric. Because each actor is assigned to one cluster at each MCMC iteration, the acceptance probability is

$$\min\left(1,\frac{f_{Y|Z_i,\delta_i,\gamma_i,\ldots}(y|Z_i^*,\delta_i^*,\gamma_i^*,\ldots)f_{\mathrm{MVN}_d(\mu_{K_i},\sigma^2_{K_i}I_d)}(Z_i^*)f_{\mathrm{N}(0,\sigma_\delta^2)}\left(\delta_i^*\right)f_{\mathrm{N}(0,\sigma_\gamma^2)}\left(\gamma_i^*\right)}{f_{Y|Z_i,\delta_i,\gamma_i,\ldots}(y|Z_i,\delta_i,\gamma_i,\ldots)f_{\mathrm{MVN}_d(\mu_{K_i},\sigma^2_{K_i}I_d)}(Z_i)f_{\mathrm{N}(0,\sigma_\delta^2)}\left(\delta_i\right)f_{\mathrm{N}(0,\sigma_\gamma^2)}\left(\gamma_i\right)}\right).$$

An independent $p$-variate normal proposal is used for $\beta$, namely

$$\beta^* \sim \mathrm{N}\left(\beta, \tau_\beta^2 I_p\right).$$

At the same time, a proposal is made to scale all the latent positions and clustering variables away from the origin by a factor drawn from the lognormal distribution:

$$h \sim \mathrm{LN}(0, \tau_Z^2),$$

and

$$\begin{aligned} Z_i^* &= hZ_i \quad i = 1, \ldots, n, \\ \mu_g^* &= h\mu_g \quad g = 1, \ldots, G, \\ \sigma_g^{2*} &= h^2\sigma_g^2 \quad g = 1, \ldots, G. \end{aligned}$$

The lognormal distribution is multiplicatively symmetric about 1 ($\forall_{h>0} f_{X^*|X}(hx|x) = f_{X^*|X}(\frac{1}{h}x|x)$), so this is still a Metropolis proposal.

In addition to that, a proposal $h_\delta|\beta^* - \beta$ is made to shift all sender effects by a random amount, and similarly $h_\gamma|h_\delta, \beta^* - \beta$ with receiver effects. The proposals are set up so that their joint distribution is multivariate normal centered at the origin, and the pairwise correlations among the three proposed jumps are negative. Thus,

$$\begin{aligned} \delta_i^* &\sim \delta_i + h_\delta \quad i = 1, \ldots, n, \\ \gamma_i^* &\sim \gamma_i + h_\gamma \quad i = 1, \ldots, n. \end{aligned}$$

The acceptance probability is

$$\min\left(1, \frac{f_{Y|\beta,Z,\delta,\gamma,\ldots}\left(y|\beta^*, Z^*, \delta^*, \gamma^*, \ldots\right) f_{\mathrm{Prior}}\left(\beta^*, \mu^*, \sigma^{2*}\right) \prod_{i=1}^n f_{\mathrm{Actor}\ i}^*}{f_{Y|\beta,Z,\delta,\gamma,\ldots}\left(y|\beta, Z, \delta, \gamma, \ldots\right) f_{\mathrm{Prior}}\left(\beta, \mu, \sigma^2\right) \prod_{i=1}^n f_{\mathrm{Actor}\ i}}\right),$$

where

$$f_{\mathrm{Prior}}(\beta, \mu, \sigma^2) = f_{\mathrm{MVN}_p(\xi,\Psi)}(\beta) \prod_{g=1}^{G} \left(f_{\mathrm{MVN}_d(0,\omega^2 I_d)}\left(\mu_g\right) f_{\alpha_Z \sigma_{0,Z}^2 \mathrm{Inv}\chi^2_{\alpha_Z}}\left(\sigma_g^2\right)\right),$$

$$f_{\mathrm{Actor}\ i} = f_{\mathrm{MVN}_d(\mu_{K_i}, \sigma^2_{K_i} I_d)}\left(Z_i\right) f_{\mathrm{N}(0,\sigma_\delta^2)}\left(\delta_i\right) f_{\mathrm{N}(0,\sigma_\gamma^2)}\left(\gamma_i\right),$$

and

$$f_{\mathrm{Actor}\ i}^* = f_{\mathrm{MVN}_d(\mu^*_{K_i}, \sigma^{2*}_{K_i} I_d)}\left(Z_i^*\right) f_{\mathrm{N}(0,\sigma_\delta^2)}\left(\delta_i^*\right) f_{\mathrm{N}(0,\sigma_\gamma^2)}\left(\gamma_i^*\right).$$

## 3.3 Identifiability of Parameters and Initialization

The likelihood is a function of the latent positions only through their distances, and so 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 (Stephens, 2000).

We use the approach of HRT to resolve these near nonidentifiabilities by post-processing the MCMC output. The approach is to find a configuration of cluster labels and positions with implied distribution close to the corresponding "true" distribution in terms of Bayes risk. See the appendix of HRT for the technical details. The post-processed actor positions are denoted by $Z_{\mathrm{MKL}}$.

For visualization purposes, posterior cluster means and variances corresponding to chosen positions are also needed. We use the full conditionals for $\mu_g$, $\sigma_g^2$, $\lambda$, and $K$ given in Section 3.2 to Gibbs-sample $\mu, \sigma^2, \lambda, K|Z_{\mathrm{MKL}}$, and we use the posterior means of $\mu|Z_{\mathrm{MKL}}$ and $\sigma^2|Z_{\mathrm{MKL}}$ as point estimates to go with $Z_{\mathrm{MKL}}$.

The proposal distribution variance parameters, $\tau_Z$, $\tau_\gamma$, $\tau_\delta$, and $\tau_\beta$, are set by the user to achieve good performance of the algorithm.

To speed convergence, we start the algorithm at an approximation to the posterior mode. Specifically:

1. Multidimensional scaling is performed on geodesic distances between the graph vertices to get initial latent space positions $Z_{\mathrm{MDS}}$ (Breiger, Boorman, and Arabie, 1975). These are then centered at the origin.

2. Model-based clustering is used to get a hard clustering $K_{\mathrm{MDS}}$ of $Z_{\mathrm{MDS}}$ (Fraley and Raftery, 2002). To improve robustness, the first time through, locations with Mahalanobis distances from the origin greater than 20 are excluded. This threshold value was found experimentally to exclude small graph components and isolates but still provide a good margin of safety for vertices containing useful information about structure. For the excluded points, $K_{\mathrm{MDS}}$ is arbitrarily assigned to the largest cluster.

3. Numerical optimization is used to find the posterior mode conditional on $K_{\mathrm{MDS}}$.

4. Steps 2 and 3 are repeated to convergence.

The proposal distribution variance parameters, $\tau_Z$, $\tau_\gamma$, $\tau_\delta$, and $\tau_\beta$, are set by the user to achieve good performance of the algorithm.

Table 1: Characteristics of Example Networks

| | Sampson's Monks | Add Health: Original | Add Health: Isolates removed | Small world: Unclustered | Small world: Clustered |
|---|---|---|---|---|---|
| directed | Yes | Yes | Yes | No | No |
| actors | 18 | 71 | 69 | 150 | 150 |
| edges | 88 | 305 | 305 | 244 | 244 |
| (density) | (0.29) | (0.061) | (0.065) | (0.022) | (0.022) |
| components (incl. isolates) | 1 | 3 | 1 | 7 | 3 |
| isolates | 0 | 2 | 0 | 0 | 0 |
| (proportion) | (0.0) | (0.028) | (0.0) | (0.0) | (0.0) |
| # actors in biggest component | 18 | 69 | 69 | 137 | 143 |
| (proportion) | (1.0) | (0.97) | (1.0) | (0.91) | (0.95) |
| (Clustering Coefficient) | – | – | – | 9.55% | 15.06% |

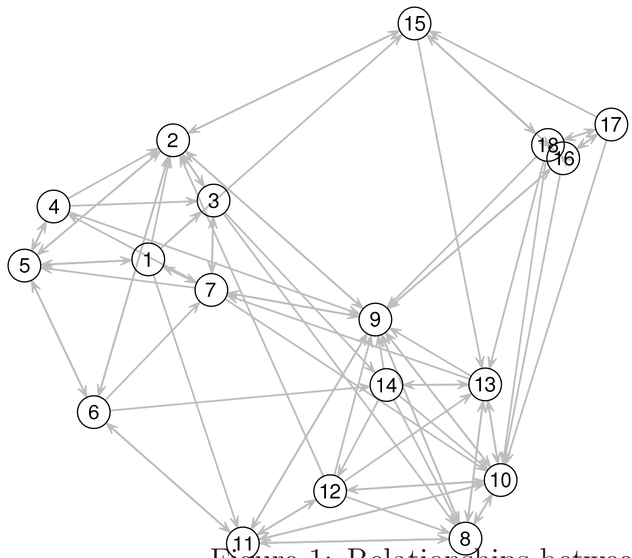
# 4 Examples

We consider four datasets summarized in Table 1. The first two have previously been analyzed using latent position and latent position cluster models, and we compare the model fits to those previously obtained. Both of these are directed. The third and fourth dataset represent two forms of "small-world" networks. Both are undirected.

## 4.1 Example 1: Liking between Monks

We consider the social relations between 18 monks in an isolated American monastery (Sampson, 1969). While resident at the monastery, Sampson 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. The minimum Kullback-Leibler estimates of the latent space positions (Shortreed et al., 2006) are presented visually in Figure 1.

The measurement process for these data imposed constraints on the monk-specific sender and receiver effects. In particular, the sender effects are limited: Sampson asked each monk to name the three others that he liked most, three times over the period of the study, so the out-degree of each monk is bounded. The dataset pools these nominations, so a tie between one monk and another exists if the first monk nominated the second as one of his top three most liked *at least once*. Thus, the number of out-ties a monk has is less a measure of the monk's sociality and more a measure of how often the monk changes his friends. On the other hand, receiver effects are interpretable.




| | | | |
|---|---|---|---|
| 1 | Ramauld | 10 | Gregory |
| 2 | Bonaventure | 11 | Hugh |
| 3 | Ambrose | 12 | Boniface |
| 4 | Berthold | 13 | Mark |
| 5 | Peter | 14 | Albert |
| 6 | Louis | 15 | Amand |
| 7 | Victor | 16 | Basil |
| 8 | Winfred | 17 | Elias |
| 9 | John | 18 | Simplicius |

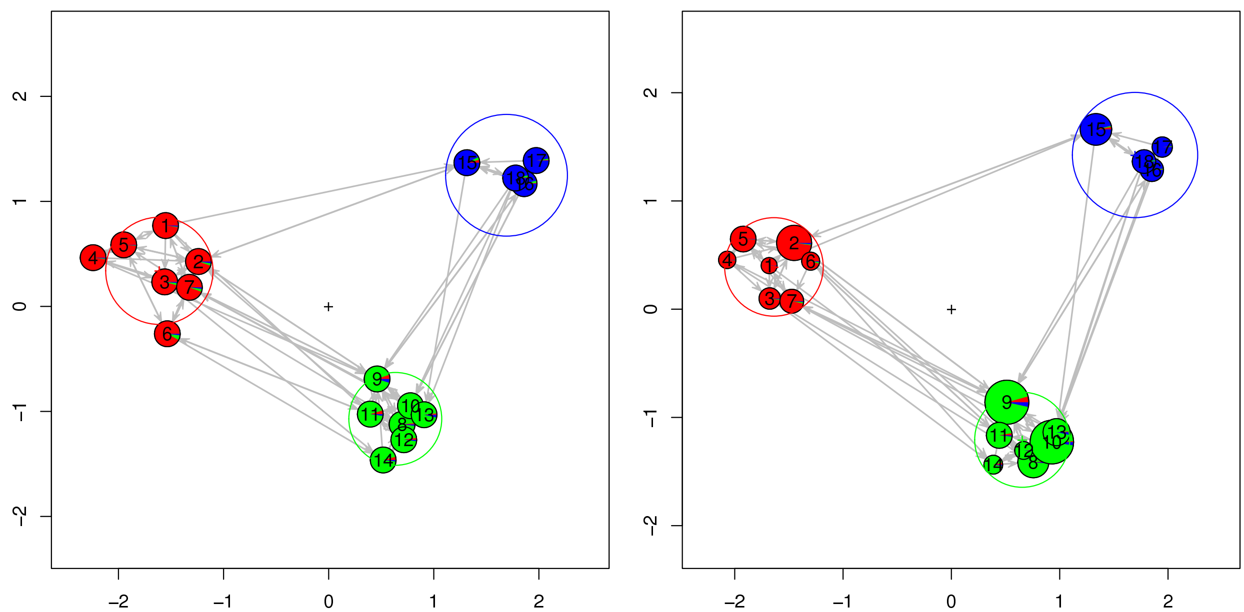
Figure 1: Relationships between monks within a monastery.

Sampson (1969) identified three main groups of monks: 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 (Sampson 1969, p. 370; White, Boorman and Breiger 1976, p.752-753.)

We fit two versions of our clustering model: a two-dimensional, three-cluster, latent space model without random effects, and one with receiver effects. We used the hyperparameter values $\nu_1 = \nu_2 = \nu_3 \approx 2.45$, $\sigma_0^2 = 0.75$, $\alpha_Z \approx 2.54$, $\sigma^2_{0,\delta} = 1.0$, $\alpha_\delta = 3$, $\sigma^2_{0,\gamma} = 1.0$, $\alpha_\gamma = 3$, and $\omega^2 = 4.5$. We fit our model using MCMC with 1,000 burn-in iterations that were discarded, and a further 20,000 iterations, of which we kept every 10th 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 fits are summarized in Figure 2. 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. The monks are well separated into the three groups and our model assigns each monk to the same group that Sampson did, with high probability. The groups from top to bottom are Loyal Opposition, Turks, followed by the combined Outcasts and Waverers. The Young Turks are also more tightly clustered than the Loyal Opposition. The performance of the model without random effects is similar to that of the latent position cluster model reported in HRT.

An interesting contrast between models with and without receiver effects is Monk #1 (Ramauld). This monk is relatively unpopular: he has out-ties to 4 of the 6 members of Loyal Opposition (as identified in Sampson's original paper), but few in-ties from anyone. In the model without receiver effects (Fig. 2a), this monk is thus pushed to the edge of the




(a) without random effects

(b) with receiver effects

Figure 2: Relationships between monks within a monastery. The ties (i.e. the data) are shown by the arrows. Panel (a) gives a latent cluster model without monk-specific random effects; panel (b) adds receiver random effects. For the latter, the area of the plotting symbol is proportional to the conditional odds ratio of a nomination for the monk due to its receiver effect. The radius of the circles is equal to the cluster standard deviations, $\sigma_g$.

Loyal Opposition group. When the receiver effects are added (Fig. 2b), this monk moves toward the center of the Loyal Opposition group because of his out-ties to them and has a small receiver effect to compensate. Thus, the receiver effect absorbs his unpopularity and the model gives him a more informative position in the social space.

### 4.2 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 (Harris et al., 2003). Here we consider a single school of 71 adolescents from grades seven through twelve. Each student 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). These data were previously analyzed by HRT, who excluded two adolescents who had no ties in the network. One visualization of the data is shown in Figure 3.

Based on the formulae given earlier, we used hyperparameters $\sigma^2_{0,Z} \approx 1.44$ and $\omega^2 \approx 17.25$. These variances are larger than for the monks' dataset in the last section, reflecting the fact that this is a larger network, and so requires more space for the latent positions. Also, $\nu_g = \alpha_Z \approx 3.39$. The results are shown in Figure 4.

We would expect the students to tend to be linked to others in their own grade, but we did not use the information about which grade a student was in when fitting the model. The extent to which estimated clusters correspond to grades can therefore be used as a partial check on the clustering procedure. The clusters did correspond quite well to the grades.

As can be seen from Figure 4(b), the estimated receiver effects for the students do not vary dramatically in this case and so the results from the model with random receiver effects do not differ greatly from those from the latent position cluster model without random effects.

### 4.3 Example 3: Scale-Free and Small-world Networks

Two features of networks that have received a great deal of recent attention are local clustering and average minimum path length, the determinants of the so-called "small world" network effect. Small-world networks can arise from power-law models of the degree distribution (Amaral et al., 2000). Small world graphs are of interest because their properties differ from purely random networks with the same density of ties. For example, the epidemic potential of small world networks is much higher than that of other sparse networks because of their relatively high connectedness (Newman, 2002).

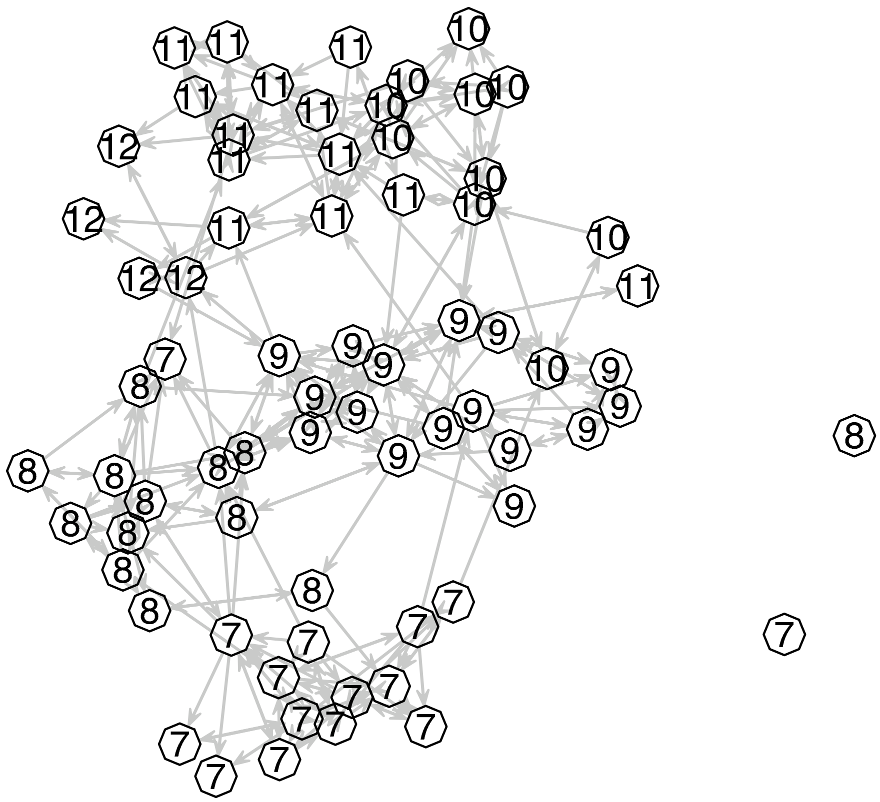
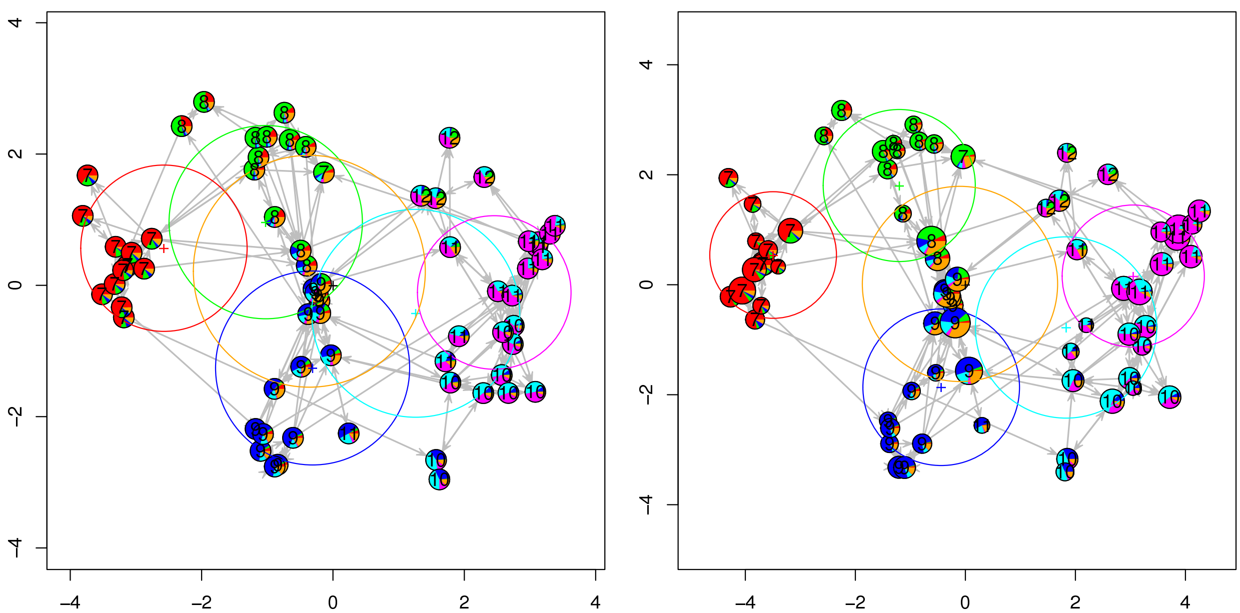


Figure 3: Visualization of the friendship network in the Add Health school. The grades of each student are shown as numbers.



(a) No random effects (b) Receiver effects

Figure 4: Estimated positions from the latent position cluster model for the adolescent health data, where the number of clusters is constrained to be 6.
Plot (a) gives a plain latent cluster model fit to the Add Health dataset (Fig. 1), and plot (b) adds a receiver effect for each student. For the latter, the area of the plotting symbol is proportional to the conditional odds ratio of a tie for its vertex due to its random receiver effect.

The measure of local clustering we use is the *clustering coefficient* defined as

$$C = \frac{3N_3^*}{N_3}$$

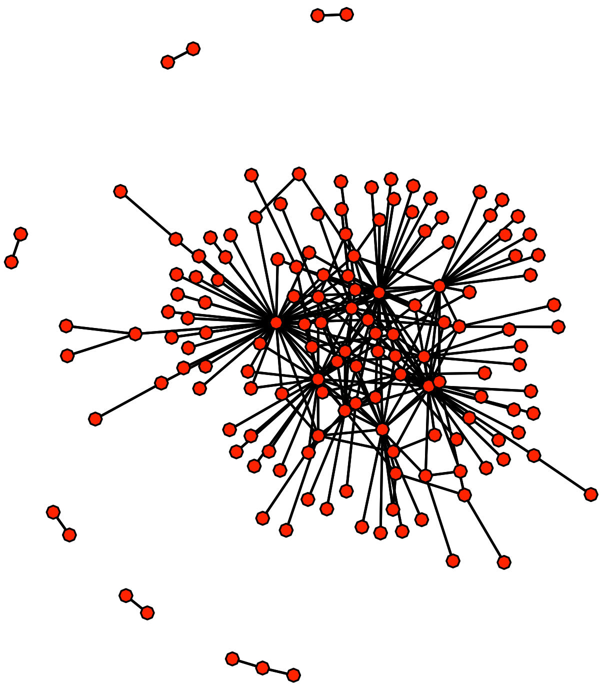
where $N_3^*$ is the number of complete subgraphs of size 3 and $N_3$ is the number of connected triples of actors, or 2-stars (Frank and Strauss, 1986; Newman, 2003). This is the proportion of complete triads in the networks out of a total number of possible triads, and is thus a measure of transitivity. Note that the term clustering is used here in a different sense than in the rest of this paper.

There has been a focus on the various "scale-free," preferential attachment and power-law models for networks, especially in the physics literature (Newman, 2003). These models assume that all networks with the same degree distribution are equally likely. Let $D$ be the degree of a randomly chosen actor in the network. We say that $P(D = k)$ has *power-law behavior* with scaling exponent $\phi > 1$ if there exist constants $c_1, c_2$, and $M$ such that $0 < c_1 \leq P(D = k)k^{\phi} \leq c_2 < \infty$ for $k > M$ (Jones and Handcock, 2003). Scale free networks are one form of small world networks, although some researchers focus on the subclass of networks with high local clustering relative to a random network with the same degree sequence (Newman, 2003).

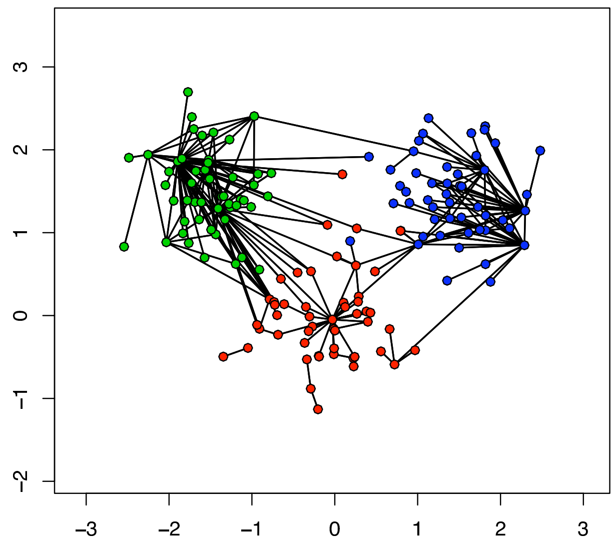
Here we consider two networks generated from small world processes. Each has 150 actors and an undirected relationship between them. They are sparse networks with density 2.2%. The first is a network following the preferential attachment model of Handcock and Jones (2004). In this model the degree sequences follow a Yule probability distribution, and the actors form ties independently given this sequence. The network generating process exhibits power-law behavior with $\phi = 2.5$. It is thus a scale-free network, and is an example of a network with no latent clustering and a very right-skewed sociality effects distribution. The technical details of this model and computational details are given in Handcock and Morris (2007). The network is visualized in Figure 5(a). Note how the high-degree actors act as "hubs" for the other actors.

The second network has the same sociality effects as the first but with latent positions drawn from the model (2) with $G = 3$ groups in $d = 2$ dimensions. The clusters are dispersed with $\mu_1 = (0, 0), \mu_2 = (-1.5, 1.5), \mu_3 = (1.5, 1.5)$ The intra-cluster standard deviation in positions is $\sigma_g = 0.2$. The network is a random draw from the Latent Cluster Random Effects Model conditional on the degree sequence of the first network. This network also satisfies preferential attachment, has power-law structure and is a small world. However it also has clustered latent positions that lead to highly clustered pattern of links.

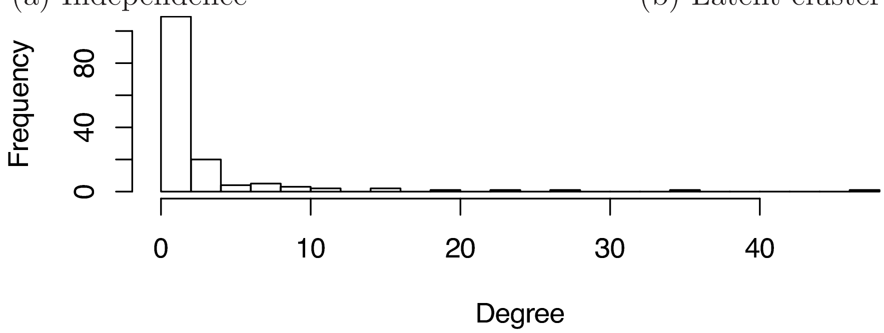
The second network is visualized in Figure 5(b). The positions are those drawn from the mixture model. Note that despite their differing appearance the two networks have the same



(a) Independence




(b) Latent cluster




(c) Degree distribution (same for both graphs)

Figure 5: Yule model networks. These 150-actor networks (a) and (b) have been generated from the Yule model, scale-free, with $\rho = 2.5$, giving the degree distribution shown in (c). In the independence network (a), the ties are independent given the degree distribution. In the latent cluster network (b), a dyad-level latent space distance term was also added. The latent space positions were generated from a bivariate normal mixture model with three components.

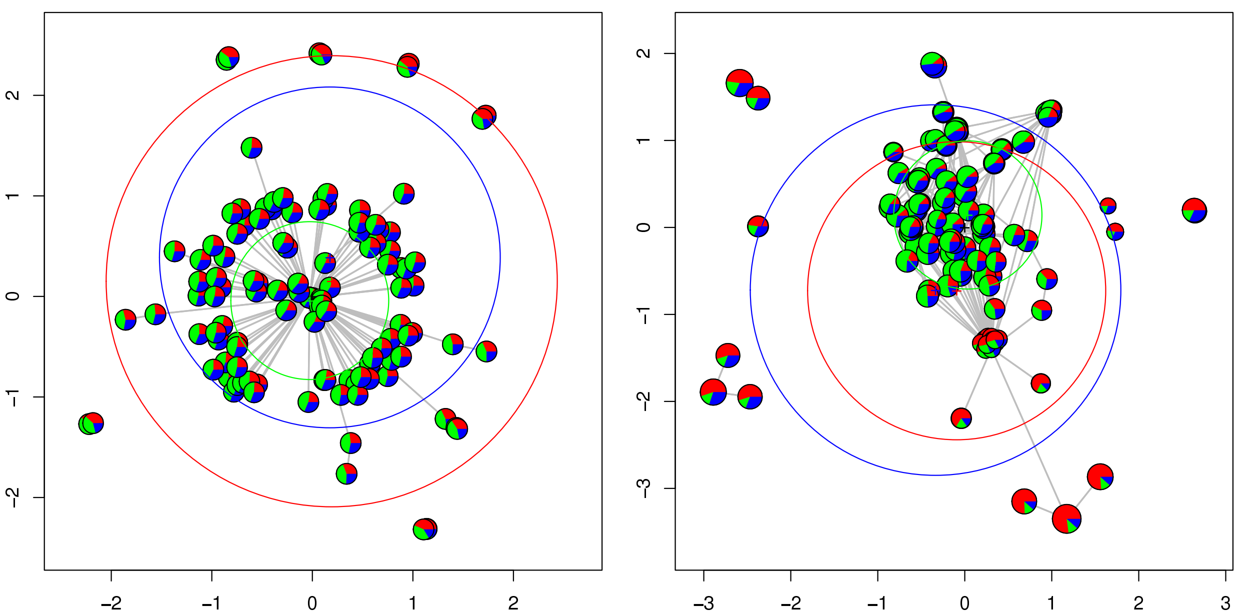
degree distribution, shown in Figure 5(c). Note the extreme right tail that is characteristic of scale-free distributions. Neither of these networks is a draw from the Latent Cluster Random Effects Model. Both have markedly non-Gaussian sociality effects, and the first does not have multiple groups.

The strength of the small world effects for the undirected networks is given by the clustering coefficient values in Table 1. The clustering coefficient for a purely random network with 150 actors and the same density as these networks is 2.04%. Thus the first network has almost five times this level of clustering. The second network has over seven times the level. Thus both networks can be classified as small worlds. Note that the latent position clustering can lead to much greater small world structure — here increasing it by over 50%. This suggests that the small world phenomenon can be induced by the presence of an unobserved social space.

We now report the results of fitting the Latent Cluster Random Effects Model to these networks. In each case, we fit two models: a latent 3-cluster model with no random effects, and a latent 3-cluster model with random sociality effect. We used the hyperparameters $\sigma^2_{0,Z} = 6.25$, $\omega^2 = 37.5$, and $\nu_1 = \nu_2 = \nu_3 = 7.07$.

The fits of the two models (with and without random sociality effects) to the unclustered small world network are shown in Figure 6. The estimated latent space positions vary very little for either model. For example, for the model with sociality effects, only about half of the estimated latent positions are a distance of more than 1 from the origin. Thus neither of the two latent space models that we fit (with and without random sociality effects) finds much evidence of distinct groups. There are no groups in the data, so both models reach the right conclusion in this case.

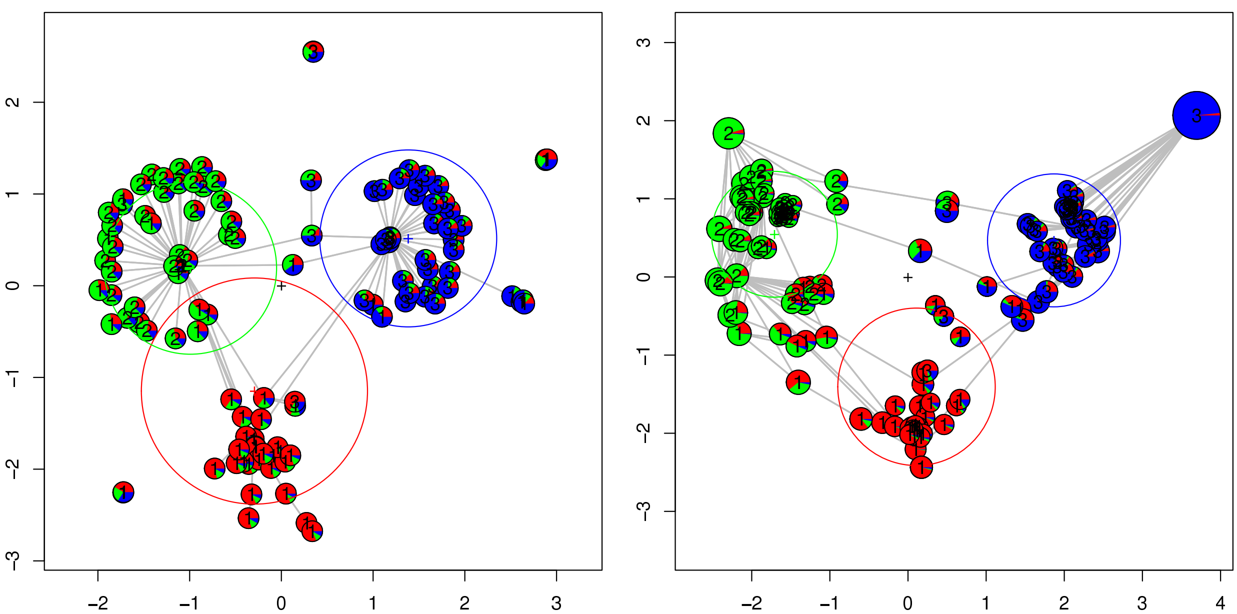
The "successful" fits of the two models to the clustered small world network are shown in Figure 7. While both models models are able to detect the distinct groups that are present in the data, we found that the model with random effects was substantially more reliable in this: in our example, 3-cluster fits of the model without random effects were only occasional (and "unsuccessful" fits look more like those fits based on the unstructured network). We believe the extremely skewed sociality levels masked the group structure under the simpler model. In contrast, the Latent Cluster Random Effects Model did find the three groups and placed them approximately in the right places and the right distances apart. For the latent cluster model with random effects, 87% of the vertices were clustered correctly (82% for the model without random effects).



(a) Latent position cluster model (no sociality effects)

(b) Latent cluster random effects model with undirected sociality effects

Figure 6: Minimum Kullback-Leibler locations from the models for the unclustered small world network in Figure 5(a). In plot (b), the area of the plotting symbol is proportional to the conditional odds ratio of a tie for its vertex, due to its random sociality effect.



(a) Latent position cluster model (no sociality effects)

(b) Latent cluster random effects model with undirected sociality effects

Figure 7: Minimum Kullback-Leibler locations from the models for the clustered small world network in Figure 5(b). In plot (b), the area of the plotting symbol is proportional to the conditional odds ratio of a tie for its vertex, due to its random sociality effect.

# 5 Discussion

We have introduced an extension to the latent space model of Hoff et al. (2002) and the latent position clustering model of HRT that also models heterogeneity in actor sociality levels by including random effects. Called the latent cluster random effects model, this gives satisfactory fits to two real datasets. Heterogeneity among actors in social networks has often been modeled using scale-free and small-world network models that assume that all networks with the same degree distribution are equally likely. With two simulated datasets we show that networks with the same degree distribution can have very different clustering behavior. Our model can capture both the degree distribution and the clustering behavior adequately.

We have focused here on binary networks, but much network data is not of this kind. For example, links may consist of counts, such as the number of phone calls two actors exchange, or amounts, such as the amount of trade between two countries. Following Hoff (2005), the latent class random effects model can be extended to deal with such situations by replacing the binary response in (1) with a different distribution, such as Poisson if the links are counts, and normal, lognormal or gamma if they are amounts.

One problem we have not addressed here is that of choosing the number of groups and the latent space dimension. This can be done by recasting the problem as one of statistical model selection and using Bayesian model selection to solve it. HRT did this for choosing the number of groups in their latent position cluster model, Oh and Raftery (2001) did it for choosing the dimension of the latent space for a related Bayesian multidimensional scaling model, and Oh and Raftery (2007) did it for choosing both the number of groups and the latent space dimension simultaneously in model-based clustering for dissimilarities. This work could be adapted and extended to the latent cluster random effects model.

We have used a Euclidean distance for our latent social space, but this is not the only possible measure on which to base the model. In particular, Hoff, Raftery, and Handcock (2002) and Hoff (2005) have proposed using an inner product, which has certain advantages.

The hyperparameter settings that we have used have varied slightly between our examples, reflecting the outcome of experimentation. It would be desirable to have a more automatic way of choosing the hyperparameters. One possibility would be to fit a logit model with node-specific effects, and then use the variances of these effects to obtain an empirical Bayes-type prior. This issue will be the focus of future research.

# References


P. D. Hoff, A. E. Raftery, M. S. Handcock, Latent space approaches to social network analysis, Journal of the American Statistical Association 97 (2002) 1090–1098.

M. S. Handcock, A. E. Raftery, J. M. Tantrum, Model-based clustering for social networks (with discussion), Journal of the Royal Statistical Society, Series A 170 (2007) 301–354.

P. D. Hoff, Bilinear mixed-effects models for dyadic data, Journal of the American Statistical Association 100 (2005) 286–295.

M. A. J. van Duijn, T. A. B. Snijders, B. H. Zijlstra, $p_2$: A random effects model with covariates for directed graphs, Statistica Neerlandica 58 (2004) 234–254.

P. D. Hoff, Random effects models for network data, in: R. Breiger, K. Carley, P. Pattison (Eds.), Dynamic Social Network Modeling and Analysis, Vol. 126, Committee on Human Factors, Board on Behavioral, Cognitive, and Sensory Sciences, National Academy Press, Washington, DC., 2003, pp. 302–322.

J. Diebolt, C. P. Robert, Bayesian estimation of finite mixture distributions, Journal of the Royal Statistical Society, Series B 56 (1994) 363–375.

M. Stephens, Dealing with label-switching in mixture models, Journal of the Royal Statistical Society, Series B, Methodological 62 (2000) 795–809.

R. L. Breiger, S. A. Boorman, P. Arabie, An algorithm for clustering relational data with application to social network analysis and comparison with multidimensional scaling, Journal of Mathematical Psychology 12 (1975) 328–383.

C. Fraley, A. E. Raftery, Model-based clustering, discriminant analysis and density estimation, Journal of the American Statistical Association 97 (2002) 611–631.

S. F. Sampson, Crisis in a cloister, Ph.D. thesis, Cornell University (1969).

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

A. E. Raftery, S. M. Lewis, Implementing MCMC, in: W. R. Gilks, D. J. Spiegelhalter, S. Richardson (Eds.), Markov Chain Monte Carlo in Practice, Chapman and Hall, London, 1996, pp. 115–130.

K. M. Harris, F. Florey, J. Tabor, P. S. Bearman, J. Jones, R. J. Udry, The national longitudinal of adolescent health: Research design [WWW document], Tech. rep., Carolina Population Center, University of North Carolina at Chapel Hill, Available at: http://www.cpc.unc.edu/projects/addhealth/design (2003).

R. J. Udry, The national longitudinal of adolescent health: (add health), waves I and II, 1994-1996; wave III, 2001-2002 [machine-readable data file and documentation], Tech. rep., Carolina Population Center, University of North Carolina at Chapel Hill (2003).

L. A. N. Amaral, A. Scala, M. Barthelemy, H. E. Stanley, Classes of small-world networks, Proceedings of the National Academy of Sciences of the United States of America 97 (2000) 11149–11152.

M. E. J. Newman, Spread of epidemic disease on networks, Physical Review E 66 (2002) art. no.–016128.

O. Frank, D. Strauss, Markov graphs, Journal of the American Statistical Association 81 (1986) 832–842.

M. E. J. Newman, The structure and function of complex networks, SIAM Review 45 (2003) 167–256.

J. H. Jones, M. S. Handcock, An assessment of preferential attachment as a mechanism for human sexual network formation, Proceedings of the Royal Society of London, B 270 (2003) 1123–1128.

M. S. Handcock, J. H. Jones, Likelihood-based inference for stochastic models of sexual network formation, Theoretical Population Biology 65 (2004) 413–422.

M. S. Handcock, M. Morris, A simple model for complex networks with arbitrary degree distribution and clustering, in: E. M. Airoldi (Ed.), Workshop on Statistical Network Analysis, ICML 2006, Pittsburgh, USA, June 29, 2006, Vol. 4503 of Lecture Notes in Computer Science, Springer, 2007, pp. 103–114.

M. S. Oh, A. E. Raftery, Bayesian multidimensional scaling and choice of dimension, Journal of the American Statistical Association 96 (2001) 1031–1044.

M. S. Oh, A. E. Raftery, Model-based clustering with dissimilarities: A Bayesian approach, Journal of Computational and Graphical Statistics 16 (2007) to appear.