Social Networks 31 (2009) 204–213

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models

Pavel N. Krivitsky [*,1,2], Mark S. Handcock [1], Adrian E. Raftery [3], Peter D. Hoff [4]

University of Washington, Seattle, United States

ARTICLE INFO

Keywords:
Bayesian inference
Latent variable
Markov chain Monte Carlo
Model-based clustering
Small world network
Scale-free network

ABSTRACT

Social network data often involve transitivity, homophily on observed attributes, community structure, and heterogeneity of actor degrees. We propose a latent cluster random effects model to represent all of these features, and we develop Bayesian inference for it. The model is applicable to both binary and non-binary network data. We illustrate the model using two real datasets: liking between monks and coreaderships between Slovenian publications. We also apply it to two simulated network datasets with very different network structure but the same highly skewed degree sequence generated from a preferential attachment process. One has transitivity and community structure while the other does not. Models based solely on degree distributions, such as scale-free, preferential attachment and power-law models, cannot distinguish between these very different situations, but the latent cluster random effects model does.

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. In this article we primarily consider binary social network data, representing presence or absence of a relationship, and count data, representing the number of times a relationship between a pair of actors was observed. The methods we develop can also be extended to accommodate for other types of relational data.

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*, in which actors cluster into groups such that ties are more dense within groups than between them. It has also been referred to as *community structure* (Newman, 2003). This can be due to social self-organization or to homophily on unobserved attributes, such as interest in the same sport, about which the analyst might not have information. A fourth feature is *degree heterogeneity*, namely the tendency of some actors to send and/or receive links more than others.

Hoff et al. (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. Inference for the model involves estimating both the characteristics of 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. This model was extended by Handcock et al. (2007) — hereafter HRT — to include model-based clustering of the latent space positions, giving a way to detect groups of actors, or so-called community structure. Hoff (2005) added random sender and receiver effects to model inhomogeneity of the actors, similar to those in the p_2 model (van Duijn et al., 2004), and described its generalized linear model formulation, applying it to non-binary data.

No model so far proposed has modeled all the four common features of social network data noted above: homophily, transitivity, community structure and heterogeneity in actor degrees. In this paper, we propose the latent cluster random effects model, which explicitly models all four features by adding the random sender and

* Corresponding author at: University of Washington, Department of Statistics, Box 354322, Seattle, WA 98195-4322, USA. Tel.: +1 206 543 8797; fax: +1 206 685 7419.
E-mail address: pavel@stat.washington.edu (P.N. Krivitsky).

[1] Supported by NIDA Grant DA012831, DoD ONR MURI award N00014–08–1–1015 and NICHD Grant HD041877.
[2] Supported by NIH Grant 8 R01EB 002137–02 and NSF Grant 0729438.
[3] Supported by NIH Grant 8 R01EB 002137–02 and NICHD Grant R01 HD054511.
[4] Supported by NSF Grant 0631531.

0378-8733/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2009.04.001

receiver or sociality effects as proposed by Hoff (2005) to HRT's latent position cluster model. We apply it to count data as well as binary network data.

In Section 2, we introduce the latent cluster random effects model. In Section 3, we describe our Bayesian method for estimating it using Markov chain Monte Carlo, as well as heuristics for prior and starting value selection. In Section 4 we illustrate the model using two real network datasets, one binary and the other consisting of counts. We also apply our method to two simulated networks with the same, highly skewed degree distribution, but very different network behaviors: one unstructured and the other exhibiting transitivity and clustering. Currently popular methods based on degree distributions cannot distinguish between these situations, but our model does.

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 $x = \{x_k\}_{k=1}^{p} \in \mathbb{R}^{n\times n\times p}$. 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, for binary data,

$$\operatorname{logit}(p(Y_{i,j}=1|Z,x,\beta)) \equiv \eta_{i,j} = \sum_{k=1}^{p}\beta_k x_{k,i,j} - \|Z_i - Z_j\|, \tag{1}$$

where $\operatorname{logit}(p) = \log(p/(1-p))$ and β denotes a vector of regression parameters to be estimated. The model accounts for transitivity, homophily on the observed attributes x, as well potential homophily on unobserved attributes via the latent space. As in HRT, we allow for clustering in the Z_i via a finite spherical multivariate normal mixture:

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

where λ_g is the probability that an actor belongs to the gth 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 position of each actor is drawn from one of G groups, where each group is centered on a different mean and dispersed with a different variance.

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 differs for directed and undirected relationships. For an undirected relationship, each actor i has a latent "sociality" denoted by δ_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 follows:

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

The sociality δ_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 that case we define both sender and receiver random effects, δ_i and γ_i, representing actor i's propensity to send and receive links, respectively. The model then becomes:

$$\eta_{i,j} = \sum_{k=1}^{p}\beta_k x_{k,i,j} - \|Z_i - Z_j\| + \delta_i + \gamma_j, \tag{4}$$

where

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

and the variances σ_δ^2 and σ_γ^2 measure heterogeneity in the 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), (2), and either (3) or (4). 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 ith actor belongs to the gth group, as is standard in Bayesian estimation of mixture models (Diebolt and Robert, 1994). We specify prior distributions as follows:

$$\beta \sim \mathrm{MVN}_p(\xi, \Psi),$$
$$\lambda \sim \mathrm{Dirichlet}(\nu),$$
$$\sigma_\delta^2 \sim \alpha_\delta \sigma_{0,\delta}^2 \mathrm{Inv}\chi^2_{\alpha_\delta},$$
$$\sigma_\gamma^2 \sim \alpha_\gamma \sigma_{0,\gamma}^2 \mathrm{Inv}\chi^2_{\alpha_\gamma},$$
$$\sigma_g^2 \overset{\text{i.i.d.}}{\sim} \alpha_Z \sigma_{0,Z}^2 \mathrm{Inv}\chi^2_{\alpha_Z} \quad g = 1,\ldots,G,$$
$$\mu_g \overset{\text{i.i.d.}}{\sim} \mathrm{MVN}_d(0, \omega^2 I_d), \quad g = 1\ldots G,$$

where ξ, Ψ, $\nu = (\nu_1,\ldots,\nu_G)$, $\sigma_{0,Z}^2$, α_Z, $\sigma_{0,\delta}^2$, α_δ, $\sigma_{0,\gamma}^2$, α_γ, and ω^2 are hyperparameters to be specified by the user.

We set ν_g equal 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 β. The other hyperparameters are not so clear-cut. 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 we use reflect these intuitions. This is discussed in more detail by Krivitsky and Handcock (2008a), and we use the hyperparameters $\sigma_{0,Z}^2 = (1/8)\sqrt[d/2]{(n/G)}$, $\alpha_Z = \sqrt{(n/G)}$, $\omega^2 = (1/4)\sqrt[d/2]{n}$, and $\nu_g = \sqrt{(n/G)}$.

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 δ_i and γ_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 which the variable being sampled is conditionally independent of, and 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:

$$\sigma_\delta^2|\delta, \ldots \sim (\alpha_\delta\sigma_{0,\delta}^2 + \sum_{i=1}^n \delta_i^2)\mathrm{Inv}\chi^2{}_{\alpha_\delta+n},$$

$$\sigma_\gamma^2|\gamma, \ldots \sim (\alpha_\gamma\sigma_{0,\gamma}^2 + \sum_{i=1}^n \gamma_i^2)\mathrm{Inv}\chi^2{}_{\alpha_\gamma+n},$$

$$\mu_g|Z, K, \sigma_g^2, \ldots \overset{\mathrm{ind}}{\sim} \mathrm{MVN}_d\left(\frac{n_g\bar{Z}_g}{n_g+\sigma_g^2/\omega^2}, \frac{\sigma_g^2}{n_g+\sigma_g^2/\omega^2}\right) \quad g = 1, \ldots, G,$$

$$\sigma_g^2|Z, K, \mu_g, \ldots \overset{\mathrm{ind}}{\sim} (\alpha_Z\sigma_{Z,0}^2 + SS_{Z_g})\mathrm{Inv}\chi^2{}_{\alpha_Z+n_g d} \quad g = 1, \ldots, G,$$

$$\lambda|K, \ldots \sim \mathrm{Dirichlet}(\nu_1 + n_1, \ldots, \nu_G + n_G),$$

$$Pr(K_i = g|\lambda, Z, \mu_g, \sigma_g^2, \ldots) = \frac{\lambda_g f_{\mathrm{MVN}_d(\mu_g,\sigma_g^2 I_d)}(Z_i)}{\sum_{k=1}^G \lambda_k f_{\mathrm{MVN}_d(\mu_k,\sigma_k^2 I_d)}(Z_i)} \quad i = 1, \ldots, n,$$

where $SS_{Z_g} = \sum_{i=1}^n 1_{K_i=g}(Z_i - \mu_g)^T(Z_i - \mu_g)$, 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

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

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

$$\delta_i^* \sim \mathrm{N}(\delta_i, \tau_\delta^2),$$
$$\gamma_i^* \sim \mathrm{N}(\gamma_i, \tau_\gamma^2).$$

The parameters Z_i^*, δ_i^*, and γ_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 have strong dependence: for example, a jump that moves an actor away from others would be associated with an increase in its random effect, to compensate.

This proposal is symmetric. Because each actor is assigned to one cluster at each MCMC iteration, the acceptance probability is

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

Once per MCMC iteration, a correlated proposal is used to jointly update β, Z, μ, σ, δ, and γ. Jumps $h_\beta \in \mathbb{R}^p$, $h_Z \in \mathbb{R}$, $h_\delta \in \mathbb{R}$, and $h_\gamma \in \mathbb{R}$ are generated from a correlated multivariate normal distribution:

$$\begin{bmatrix} h_\beta \\ h_Z \\ h_\delta \\ h_\gamma \end{bmatrix} \sim \mathrm{MVN}_{p+1+1+1}(0, \tau_{\beta,Z,\delta,\gamma}),$$

and updates are proposed as follows:

$$\beta^* = \beta + h_\beta,$$
$$Z_i^* = \exp(h_Z)Z_i \quad i = 1, \ldots, n,$$
$$\mu_g^* = \exp(h_Z)\mu_g \quad g = 1, \ldots, G,$$
$$\sigma_g^{2*} = \exp(2h_Z)\sigma_g^2 \quad g = 1, \ldots, G,$$
$$\delta_i^* = \delta_i + h_\delta \quad i = 1, \ldots, n,$$
$$\gamma_i^* = \gamma_i + h_\gamma \quad i = 1, \ldots, n.$$

This proposal accommodates expected posterior dependencies. The proposals to scale latent space positions, means, and variances are not symmetric in the Metropolis sense, but can be viewed as symmetric proposals on the log of the magnitudes of these variables expressed in polar coordinates. It can be shown that the acceptance ratio should be multiplied by h_Z^{nd} for latent space positions, h_Z^{Gd} for latent cluster means, and h_Z^{2G} for latent cluster variances.

The acceptance probability is thus

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

with

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

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

$$f_{Y|\beta,Z,\delta,\gamma,\ldots}^* = f_{Y|\beta,Z,\delta,\gamma,\ldots}(y|\beta^*, Z^*, \delta^*, \gamma^*, \ldots)$$
$$f_{Y|\beta,Z,\delta,\gamma,\ldots} = f_{Y|\beta,Z,\delta,\gamma,\ldots}(y|\beta, Z, \delta, \gamma, \ldots),$$

and

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

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

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 non-identifiabilities 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. This is done by minimizing the Kullback–Leibler divergence between the distribution of networks predicted by the configuration of positions and the posterior predicted distribution of networks. These are called *Minimum Kullback–Leibler* (MKL) positions (Shortreed et al., 2006). The post-processed actor positions are denoted by Z_{MKL}.

A further source of non-identifiability is that adding a constant to all of the actors' sender, receiver, or sociality effects and subtracting it from β_0 (where the corresponding covariate matrix x_0 is a matrix of ones, controlling the overall density of the network) also preserves the likelihood. While the prior distributions resolve this non-identifiability, we found that it resulted in slow mixing in our MCMC sampling, and addressed it using the correlated proposal described above.

For visualization purposes, posterior cluster means and variances corresponding to chosen positions are also needed. We use

Table 1
Characteristics of example networks.

	Sampson's Monks	Unclustered simulated network	Clustered simulated network	Slovenian publications
Directed	Yes	No	No	No
Data	Binary	Binary	Binary	Count
Actors	18	150	150	124
Density/mean	0.29	0.022	0.022	85.74
(non-0 edges)	(88)	(244)	(244)	(5972)

the full conditionals for μ_g, σ_g^2, λ, and K given in Section 3.2 to Gibbs-sample $\mu, \sigma^2, \lambda, K|Z_{\text{MKL}}$, and we use the posterior means of $\mu|Z_{\text{MKL}}$ and $\sigma^2|Z_{\text{MKL}}$ as point estimates to go with Z_{MKL}.

The proposal distribution variance parameters, $\tau_Z, \tau_\gamma, \tau_\delta, \tau_{\beta,Z,\delta,\gamma}$, are set by the user to achieve good performance of the algorithm. In practice, adaptive sampling is used (Krivitsky and Handcock, 2008a).

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_{MDS} (Breiger et al., 1975). These are then centered at the origin.
(2) Model-based clustering is used to get a hard clustering K_{MDS} of Z_{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_{MDS} is arbitrarily assigned to the largest cluster.
(3) Numerical optimization is used to find the posterior mode conditional on K_{MDS}.
(4) Steps 2 and 3 are repeated to convergence.

We implemented the algorithms in the open-source package, `latentnet` (Krivitsky and Handcock, 2008b) written for `R` (R Development Core Team, 2008). It was used to analyze the examples in this paper.

4. Examples

We consider four datasets, summarized in Table 1. The first, liking among monks in a monastery, has previously been analyzed using latent position and latent position cluster models, and we compare the model fit to those previously obtained. The second and third datasets are simulated. Both have the same degree distribution, but one has both transitivity and clustering, while the other has neither. The last dataset is a network of Slovenian newspapers, magazines, and journals, with a count, for each pair of these publications, of Slovenians surveyed who reported reading both of them. This allows us to apply this family of models to non-binary data, and provides an example of a situation where heterogeneity of actors is better modeled using fixed effects.

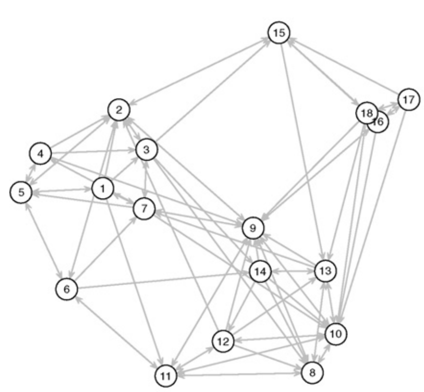
4.1. *Example 1: Liking between monks*

Our first example is the Sampson's Monks dataset: relations of "liking" among 18 monks in a monastery (Sampson, 1969). The network analyzed has a directed edge between two monks if the sender monk ranked the receiver monk in the top three monks for positive affection in any of the three interviews given over a twelve month period. The sociogram of this dataset is shown in Fig. 1.

The measurement process for these data imposed constraints on the monk-specific sender 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, the in-ties were not constrained, so a monk's receiver effect can be interpreted as the popularity of the monk, to the extent that it is reflected by how many others nominate him as a friend.

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 et al., 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



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

Fig. 1. Relationships among monks within a monastery and their affiliations as identified by Sampson: Young (T)urks, (L)oyal Opposition, (O)utcasts, and (W)averers.

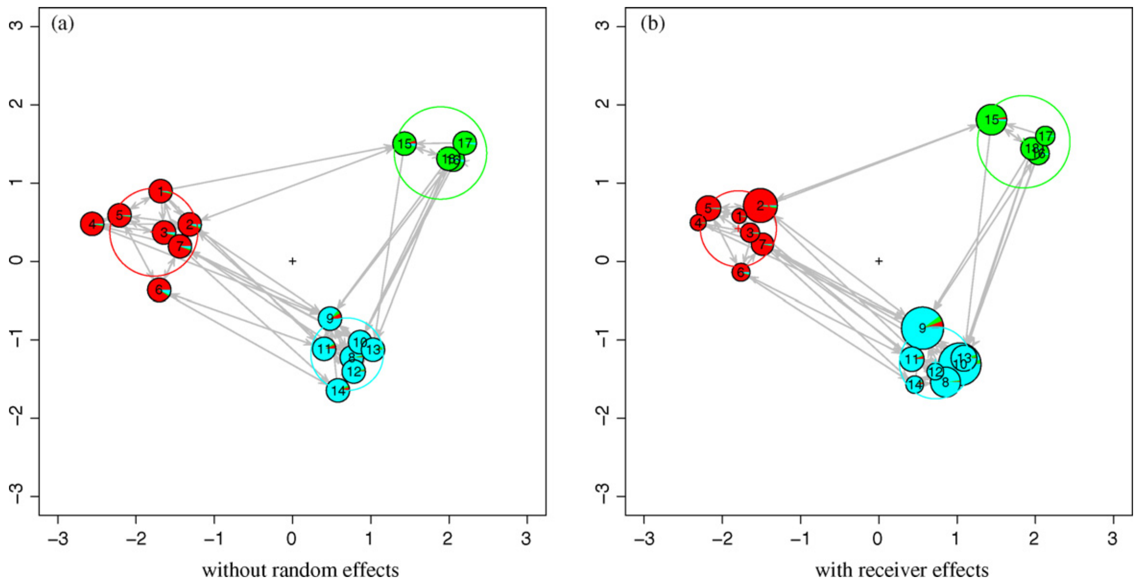


Fig. 2. Minimum Kullback–Leibler estimates of positions in the social space of monks within a monastery. Panel (a) gives estimates from a latent cluster model without monk-specific random effects; panel (b) adds receiver random effects. For the latter, the area of the pie chart is proportional to the conditional odds ratio of a nomination for the monk due to his receiver effect (also estimated using MKL), and the pie chart represents the proportions of the MCMC draws assigning each monk to each cluster. The radii of the unfilled circles are equal to the cluster standard deviations, σ_g, conditional on the MKL point estimates.

with receiver effects. In accordance with the heuristic described in Section 3.1, the hyperparameter values used were $\nu_1 = \nu_2 = \nu_3 \approx 2.45$, $\sigma_0^2 = 0.75$, $\alpha_Z \approx 2.54$, $\sigma_{0,\delta}^2 = 1.0$, $\alpha_\delta = 3$, $\sigma_{0,\gamma}^2 = 1.0$, $\alpha_\gamma = 3$, and $\omega^2 = 4.5$. The MCMC algorithm described was run, with 10,000 burn-in iterations that were discarded, and a further 40,000 iterations, of which we kept every 10th value. Visual inspection of trace plots and more formal assessments of convergence (e.g. those proposed by Raftery and Lewis (1996)), indicated that the sampling converged and that the number of iterations we used was sufficient.

The fits are summarized in Fig. 2. From the plots, the monks are well separated into the three groups and our model assigns each monk to the same group that Sampson did: all monks of Loyal Opposition (and two of the Waverers) are reliably assigned to the "Red" cluster, all the Young Turks to the "Blue" cluster, and all the Outcasts (and one Waverer) to the "Green" cluster. The Young Turks are also more tightly clustered than the Loyal Opposition. (The posterior means of the variances for their clusters are, respectively, 0.716 and 1.09 for the model without receiver effects and 0.716 and 0.968 for the model with receiver effects.)

An interesting contrast between models with and without receiver effects is Monk #1 (Ramauld, a Waverer). 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 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 low receiver effect to compensate. Thus, his position is more determined by his relations to other monks than by his overall unpopularity, which is accounted for by the receiver effect.

4.1.1. *Performance of the estimators: a simulation study*

We use the results from fitting the latent cluster receiver effects model to verify that the model and our implementation of it are able to recover the latent positions. Among the 18 monks, there are only $18 \times 17 = 306$ directed dyads — binary observations — and the latent cluster receiver effects model of dimension 2 has 55 continuous parameters in the likelihood, so in order to test whether the model is able to recover latent space positions with any accuracy, we must artificially increase the precision of the estimates. To do this, we simulated 200 networks based on 200 draws of parameter configurations from the posterior distribution of the latent cluster random effects model, and, for every ordered pair of monks, counted the number of simulated networks in which a tie on that pair was observed. We then fit a latent cluster receiver effects model with binomial response with 200 trials.

The results are summarized in Fig. 3. The latent space positions from the fit based on the summed network are very close to those from the original fit (average Euclidean distance between their MKL estimates for each actor is 0.18) as are the receiver effects.

4.2. *Example 2: A preferential attachment network with and without transitivity and community structure*

There has been a focus on scale-free, preferential attachment and power-law models for networks, especially in the physics literature (Newman, 2003). It is common in these models to assume that all networks with the same degree distribution are equally likely. As a result, methods based on these models cannot distinguish between networks that have the same degree distribution but whose network structure differs in other ways. The purpose of this simulated example is to show that our methods can make these distinctions.

We consider two networks with the same degree sequence generated via a preferential attachment process (Handcock and Jones, 2004). The first one does not exhibit either transitivity or community structure while the second has both.

Each of our simulated networks has 150 actors and an undirected relationship between them. They are sparse networks with density 0.022. The first network was simulated from the preferential attachment model of Handcock and Jones (2004) using the methods of Handcock and Morris (2007). In this model the degree sequences follow a Yule probability distribution, with $\rho = 2.5$, and the actors form ties independently given this sequence. The network generating process exhibits power-law behavior with scaling exponent 2.5. It is thus a scale-free network with a very

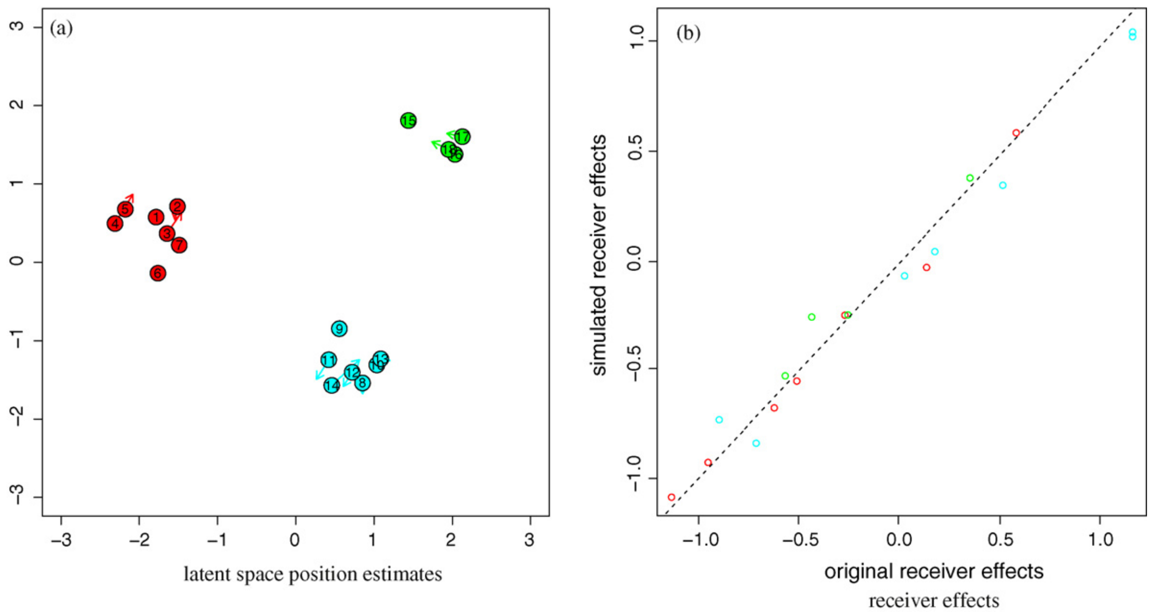


Fig. 3. Recovery of latent space positions and receiver effects from data simulated from the posterior of the latent cluster random effects model fit to Sampson's Monks. Panel (a) gives the change from the MKL estimates of latent space positions based on the original Sampson's Monks dataset to the MKL latent space positions based on the simulated data (rotated and centered). Panel (b) shows an actor's MKL receiver effect based on the Sampson's Monks fit plotted against the MKL receiver effect based on the simulated data.

right-skewed degree distribution, and exhibits no transitivity or clustering. The degree sequence is generated from the Yule distribution and the network generated using an exponential-family random graph model conditional on that degree sequence using `statnet`(Handcock et al., 2003b). The network is visualized in Fig. 4(a). Note how the high-degree actors act as "hubs" for the other actors.

The second network has the same degree distribution 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 Model conditional on the degree sequence of the first network. This network also has a power-law degree distribution. Unlike the first network, it exhibits transitivity and has clustered latent positions that lead to highly clustered pattern of links.

The two networks are shown in Fig. 4. They look very different, but they have the same degree distribution, shown in Fig. 4(c). Note the extreme right tail that is characteristic of scale-free distributions.

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 effects, both of these with 2-dimensional latent spaces ($Z_i \in \mathbb{R}^2$). We used the hyperparameters $\sigma^2_{0,Z} = 6.25$, $\omega^2 = 37.5$, and $\nu_1 = \nu_2 = \nu_3 = 7.07$, based on the heuristic described in Section 3.1.

The fits of the two models (without and with random sociality effects) to the unstructured Yule network are shown in Fig. 5. The estimated latent space positions vary very little for either model, with a possible exception of the few very high-degree nodes, and, more importantly, the estimated cluster distributions overlap almost completely. Thus, neither of the two latent space models that we fit finds much evidence of structure or distinct groups. And in fact there are no groups in the data, so both models reach the right conclusion in this case.

The fits of the two models to the clustered network are shown in Fig. 6. Both models were able to detect the distinct groups that are present in the data—the "Red" cluster is mostly group 1, "Green" is group 2, and "Blue" is group 3.

To evaluate the quality of the clustering, we use a pairwise metric similar to the Fowlkes–Mallows Index (Fowlkes and Mallows, 1983): given that two nodes drawn at random are from the same true cluster, what is the probability that the clustering algorithm assigned them to the same cluster? When using hard clustering (by assigning a node to the cluster to which the plurality of draws from the posterior assign it) this probability is 80% for the model with random sociality effects, and 78% for the model without. However, looking at the soft clustering, where the metric defined above is averaged over the posterior distribution, the difference is more pronounced: 73% for the model with sociality effects and 65% without. Both models identified the clusters of actors in the data quite well, but the random effects model did so more robustly.

Also of note is the difference in the patterns of estimated latent positions. The model without random effects gives the "Green" and "Blue" clusters a hub-and-spokes shape: a few high-degree nodes in the middle, with many low-degree nodes in a ring around them, attracted by their ties to the "hub" nodes, but repelled by their lack of ties to each other. On the other hand, the model with random sociality effects addresses this by giving the high-degree nodes a high sociality effect, low degree nodes low sociality effects, and allowing them to be positioned together, reflecting structure adjusted for degree.

This example illustrates that networks with the same degree distribution can have very different network behavior. Methods based on degree distributions, such as those based on scale-free, preferential attachment and power-law models (Newman, 2003), cannot detect these differences. However, our model clearly distinguished between networks with and without transitivity and clustering behavior.

4.3. Example 3: Slovenian newspaper and magazine coreaderships

In 1999 and 2000, CATI Center Ljubljana conducted a survey, asking over 100,000 people which newspapers, magazines, and other publications they read, producing a 2-mode, or affilia-

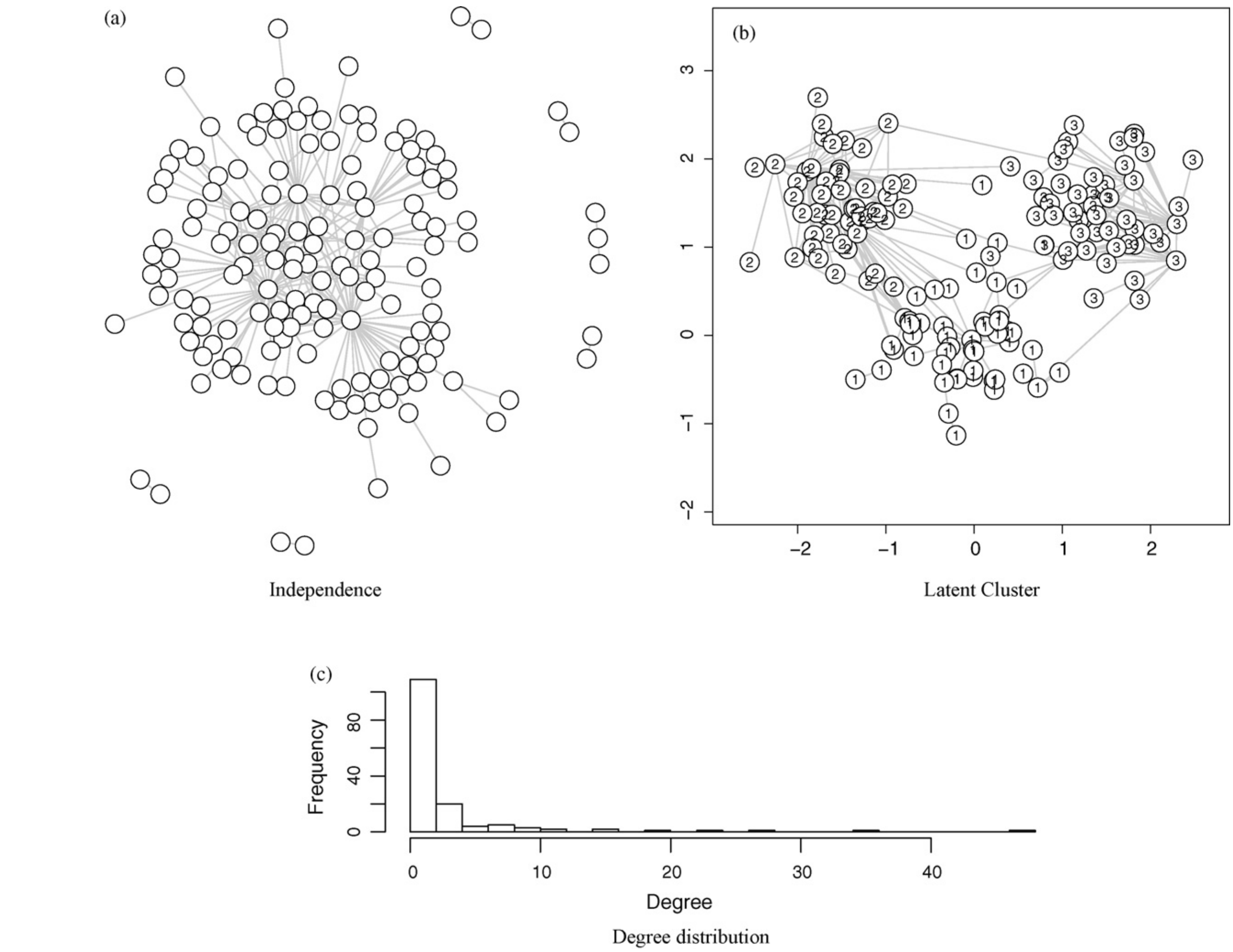


Fig. 4. Two simulated networks, each with 150 actors and the same degree distribution shown in (c). (a) Yule network (with no transitivity or clustering); (b) Latent Cluster network, where the labels 1–3 give the true cluster memberships.

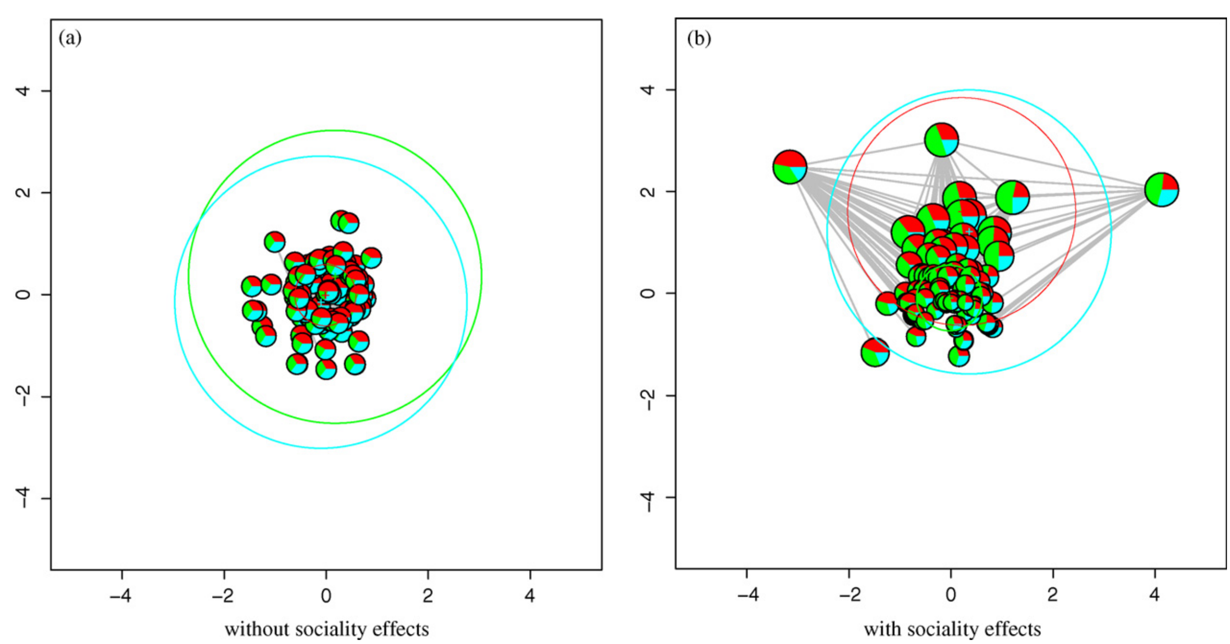


Fig. 5. Minimum Kullback–Leibler locations from the models for the unclustered network in Fig. 4(a). In plot (b), the area of the plotting symbol is proportional to the conditional odds ratio of a tie for its vertex, due to its random sociality effect. For the purposes of visualization, we limit the radii of the plotting symbols so that those vertices with sociality effects of less than −1 are plotted with the symbol size corresponding to −1, and those with sociality effects of greater than +1 are plotted with the symbol size corresponding to +1.

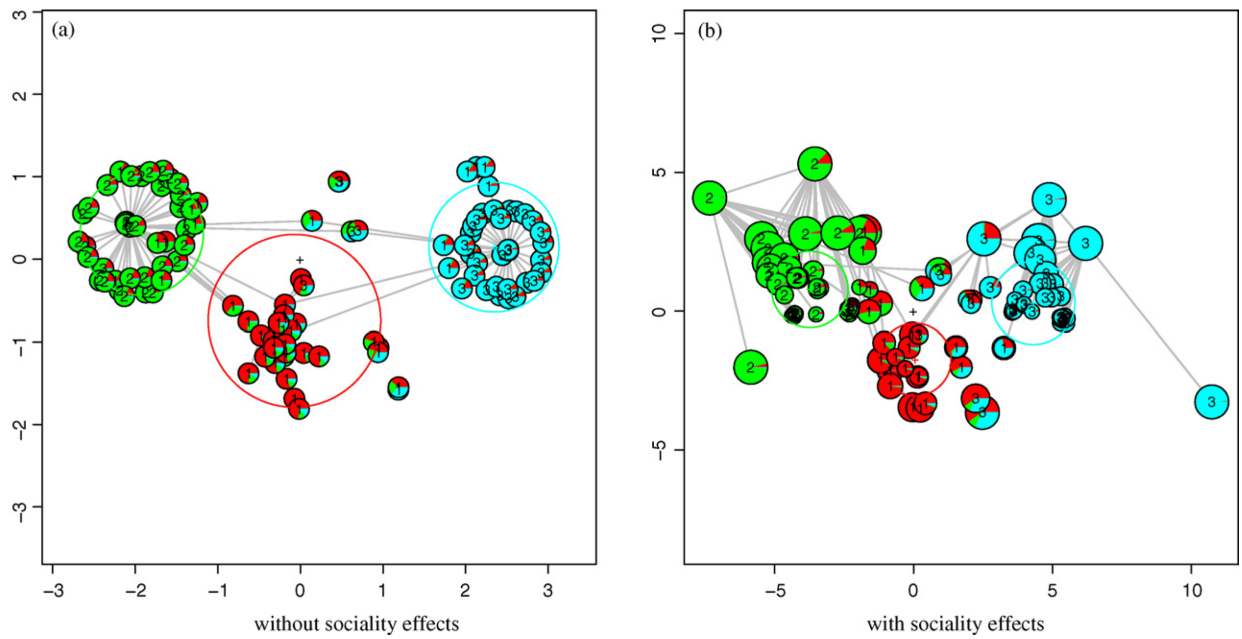


Fig. 6. Minimum Kullback–Leibler locations from the models for the clustered network in Fig. 4(b). Note the differing plot scales. In plot (b), the area of the plotting symbol is proportional to the conditional odds ratio of a tie for its vertex, due to its random sociality effect. For the purposes of visualization, we limit the radii of the plotting symbols so that those vertices with sociality effects of less than −1 are plotted with the symbol size corresponding to −1, and those with sociality effects of greater than +1 are plotted with the symbol size corresponding to +1. The numbers 1–3 give the original cluster assignments.

tion, network representing which readers read which publications. These data were then compiled into a 1-mode, undirected network of publications as follows: for a pair of publications, the number of respondents who read both was counted, producing a weighted network of "coreaderships". The dataset also breaks the publications down into 14 groups by type, topic, and audience: daily newspapers, weekly news and analysis, computers, business, home improvement, fashion, subjects traditionally of interest to men, subjects traditionally of interest to women, special interest (a catch-all category), women's, TV guides, regional, teen, and free. For each publication, the total number of respondents who reported reading it was also recorded. These data are available as a Pajek dataset "Revije" or "Journals" (Batagelj and Mrvar, 2006). This has previously been analyzed in de Nooy et al. (2005).

We analyze this network to illustrate the application of our model to non-binary data, as well as an example of a situation where a fixed covariate effect can be used in conjunction with a latent cluster model.

The coreadership for each pair of publications is a count of events (i.e. the respondent reporting that he or she reads that pair of publications) with a huge number of potential events (over 100,000). Those events (respondents) are independent, so it would be reasonable to approximate the distribution of counts as Poisson. The model is as follows:

$$Y_{i,j}|\mu_{i,j} \sim \text{Poisson}(\mu_{i,j}) \tag{5}$$

$$\log(\mu_{i,j}) = \eta_{i,j} = \beta_0 - \|Z_i - Z_j\|. \tag{6}$$

Here, the latent position Z_i of a publication i can be interpreted as its position in a space of publication appeals and interest groups, with clusters becoming those of target audience types.

Publication-specific random sociality effects (i.e. δ_i and δ_j in $\eta_{i,j} = \beta_0 - \|Z_i - Z_j\| + \delta_i + \delta_j$) could represent the overall popularity of the publication: a more popular publication would have more coreaderships. However, the overall popularity of the publication was observed directly: the number of readers of each publication was tallied. Thus, rather than using random sociality effects, we use fixed readership effects:

$$\eta_{i,j} = \beta_0 + \beta_1 x_{1,i,j} - \|Z_i - Z_j\|,$$

where $x_{1,i,j}$ is a function of the number of publication readers. We expect the number of coreaderships of a given pair of publications to be approximately proportional to their readerships, so we use $x_{1,i,j} = \log(r_i) + \log(r_j)$, where r is a vector of publication total reader counts, and set the prior mean of β_1 (which we called ξ_1) to 1 to reflect this prior information.

This resembles somewhat the association model of Goodman (1985) but the specification of the model is not the same. The idea of scores for the categories that are estimated from the data is also present in Goodman's approach. However, this network cannot be considered as a contingency table, because each respondent in the original survey could name as many publications as he or she wanted, incrementing multiple coreadership counts at once.

To explore the strength of homophily exhibited by the categories, we fit a non-latent-space quasi-independence model of the following form:

$$\eta_{i,j} = \beta_0 + \beta_1(\log(r_i) + \log(r_j)) + \sum_{k=1}^{14} \beta_{1+k} 1_{c_i=k \wedge c_j=k},$$

where $\eta_{i,j}$ are defined as in (5) and c_i and c_j are defined as the categories of publications i and j, respectively. Under this model, if two publications both belong to category k, their expected coreadership is elevated by the multiple $e^{\beta_{k+1}}$ compared to the coreadership where they do not belong to the same category. Hence a positive β_{k+1} indicates that publications in category k have disproportionately high coreadership, and a negative β_{k+1} indicates that they have a disproportionately low coreadership.

We show the maximum likelihood estimates in Table 2. The estimated coefficient of $\log(r_i) + \log(r_j)$ is very close to 1, confirming our expectation that the coreaderships are approximately proportional to the readerships of the publications involved. The signs and magnitudes of the coefficients of the homophily terms can inform our expectations of what categories will be successfully clustered.

Table 2
Coreadership network: differential homophily on categories.

Term	Coef.	Estimate	Std. Err.
Edges	β_0	−11.480	0.014
log(readership)	β_1	1.008	0.001
Both publications categorized. . .			
Business	β_2	+0.863	0.014
Computers	β_3	+2.226	0.019
Fashion	β_4	+3.325	0.053
Free	β_5	+0.798	0.248
Home Improvement	β_6	−0.072	0.043
Men's Interest	β_7	+1.310	0.031
Regional	β_8	−2.331	0.107
Special Interest	β_9	+0.559	0.022
Teen	β_{10}	+1.554	0.022
TV Guides	β_{11}	−0.281	0.013
Weekly News	β_{12}	+0.152	0.011
Women's	β_{13}	+0.416	0.006
Women's Interest	β_{14}	+1.540	0.030
Daily News	β_{15}	−0.696	0.008

Table 4
Publications assigned to Cluster 2, ordered by firmness of assignment.

Title		Category	Topics
Slovene	English		
Ribič	*Fisherman*	Men's Interest	fishing
Lovec	*Hunter*	Men's Interest	hunting
Gasilec	*Firefighter*	Special Interest	firefighting
Vzajemnost	*Reciprocity*	Special Interest	health, retirement
Družina	*Family*	Weekly News	Catholicism
Ognjišče	*Hearth*	Special Interest	Catholicism
O konjih	*About Horses*	Special Interest	horses, equestrianism
Moj mali svet	*My Little World*	Special Interest	gardening, decoration
Slovenske bradze	*Slovene Furrow*	Special Interest	agriculture
Kmetovalec	*Farmer*	Business	agriculture
Kmečki glas	*Rural Voice*	Business	agriculture

Note that only those publications assigned to Cluster 2 in the majority of posterior draws, rather than plurality, are listed.

Using a three-dimensional latent space allowed the model to detect a fairly consistent clustering with up to 5 clusters, which successfully separates those publication categories that had within-category homophily, such that publications within that category had greater-than-expected coreader counts with each other. The clustering also detected at least one reader demographic not identified by the categories. We found that a two-dimensional latent space could not adequately represent the structure in the data, and produced no clusters.

The most informative fit in three dimensions was obtained using a 6-cluster model. One of the clusters did not have the plurality of MCMC draws assign any publications to it, after dealing with label-switching as recommended by Stephens (2000), but including it seemed to facilitate mixing, as fitting a model with 5 clusters resulted in 4 non-empty clusters. The clustering is not very strong, in the sense that for many of the publications, no single cluster has a clear majority of the posterior mass for a publication assigned to it. However, it does detect some of the categories.

The cross-tabulation between clustering and known categories is given in Table 3. All the publications in each of the categories with very high homophily coefficients in Table 2 (Computers and Fashion) were assigned to the same clusters, and the majority of the posterior mass would put a pair of publications within one of these categories into the same cluster. Men's Interest and Teen magazines also had high coefficients, and tended to be sorted into the same clusters, though not as consistently. Computer magazines and Men's Interest magazines consistently share a cluster, possibly reflecting the demographics of computer professionals and hobbyists in Slovenia. On the other hand, Women's Interest magazines were not sorted into the same clusters to the same extent, despite their high coefficient. Groups of publications with small or negative homophily coefficients tended to be spread out across clusters.

Furthermore, 11 of 13 publications categorized as Business were assigned to Cluster 4, with the remaining 2 assigned to Cluster 2, and 10 of 12 magazines categorized as Men's Interests were assigned to Cluster 5, with the remaining 2 also assigned to Cluster 2. Further examination of these publications, using Google Search (2009) and Google Translate (2009) services, showed the two Business magazines, *Kmečki glas* (*Rural Voice*) and *Kmetovalec* (*Farmer*), to be about agriculture, and the two "Men's Interest" magazines, *Ribič* (*Fisherman*) and *Lovec* (*Hunter*) to be, respectively, about fishing and hunting. These and other publications with the majority of their posterior mass in Cluster 2 are listed in Table 4. They are ordered by their posterior probability of being assigned to this cluster. These publications include religious, gardening, and retiree magazines, which suggests that this cluster may capture a rural, older, more conservative demographic.

In this example actor degree effects are observed directly rather than being inferred, and are modeled as fixed rather than random effects, providing an example of combining covariate effects with latent space network models, and usefulness of this class of models for detecting clusters in networks with weighted edges is also demonstrated. While this is a network of publications, the network's clusters appear to represent reader demographics more than publication types: indeed the clustering algorithm appears to have discovered a demographic which was not included in the categories of the dataset. We also found that in this situation, the sampling

Table 3
Clustering versus categories.

Category	Homophily coefficient	Cluster					Quality metric	
		1	2	3	4	5	Hard	Soft
Business	0–1		2		11		74%	45%
Computers	> 2					8	100%	59%
Fashion	> 2				4		100%	60%
Free	0–1	2	3	2	2	4	22%	22%
Home Improvement	< 0		1		3		62%	28%
Men's Interest	1–2		2			10	72%	41%
Regional	< 0	5	2	1			47%	28%
Special Interest	0–1	3	7	2	8	3	26%	21%
Teen	1–2	5					100%	46%
TV Guides	< 0	2	1	1		1	28%	20%
Weekly News	< 0	2	2		2		28%	18%
Women's	0–1	3	2		2		35%	25%
Women's Interest	1–2	3			5		43%	25%
Daily News	< 0	2	1		1	1	28%	17%

Here, we use the same measure of quality of clustering as in the previous example, broken down by category in the dataset.

algorithm may effectively use one of the clusters to facilitate detecting the others.

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 or with fixed covariates. We found this to give satisfactory fits to two real network datasets, one with binary data consisting of the presence or absence of relationships, and one with count data. We also applied our method to two simulated networks with the same, highly skewed degree distribution, but very different network behavior: one with transitivity and clustering and other without. Currently popular methods based on the degree distribution only cannot distinguish between such very different kinds of networks, but our model is able to do so.

For directed data we have limited ourselves to modeling the two random effects of each individual as uncorrelated. Hoff (2005) and van Duijn et al. (2004) modeled the sender and receiver effects for the same individual as correlated, using a bivariate normal distribution with a Wishart prior. This is a natural extension to the latent cluster random effects model.

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 so for choosing the dimension of the latent space for a related Bayesian multidimensional scaling model, and Oh and Raftery (2007) did this 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. In the case of latent cluster models with spherical multivariate normal clusters, using Euclidean distance to determine tie probabilities has the advantage of consistency with the clustering process: the probability of a particular node belonging to a particular cluster is a function of Euclidean distances from the center of each cluster to that node. However, this is not the only possible measure on which to base the model. In particular, Hoff et al. (2002) and Hoff (2005) used an inner product, which has certain advantages. Schweinberger and Snijders (2003) proposed using an ultrametric distance.

While we provide a reasonable heuristic for our choice of hyperparameters, the heuristic itself is a result of experimentation, and it would be desirable to have a more principled 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.

Acknowledgements

Krivitsky thanks Marko Pahor for his help in identifying Slovenian publications. The authors thank Carter T. Butts, David R. Hunter, Steven M. Goodreau, and Martina Morris for helpful discussions. 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.

References

Batagelj, V., Mrvar, A., 2006. Pajek Datasets [WWW document], Available at: http://vlado.fmf.uni-lj.si/pub/networks/data/.
Breiger, R.L., Boorman, S.A., Arabie, P., 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.
Diebolt, J., Robert, C.P., 1994. Bayesian estimation of finite mixture distributions. Journal of the Royal Statistical Society, Series B 56, 363–375.
Fowlkes, E.B., Mallows, C.L., 1983. A method for comparing two hierarchical clusterings. Journal of the American Statistical Association 78, 553–569.
Fraley, C., Raftery, A.E., 2002. Model-based clustering, discriminant analysis and density estimation. Journal of the American Statistical Association 97, 611–631.
Goodman, L.A., 1985. The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries. The Annals of Statistics 13, 10–69.
Google Search, Google Inc., Mountain View, CA, Available at: http://www.google.com.
Google Translate, Google Inc., Mountain View, CA, Available at: http://translate.google.com.
Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Morris, M., 2003. `statnet`: Software Tools for the Statistical Modeling of Network Data, version 2.0, Statnet Project, Seattle, WA, Available at: http://statnetproject.org, http://CRAN.R-project.org/package=statnet.
Handcock, M.S., Jones, J.H., 2004. Likelihood-based inference for stochastic models of sexual network formation. Theoretical Population Biology 65, 413–422.
Handcock, M.S., Morris, M., 2007. A simple model for complex networks with arbitrary degree distribution and clustering, In: Airoldi, E.M. (Ed.), Workshop on Statistical Network Analysis, ICML 2006. Pittsburgh, USA, June 29, 2006. Vol. 4503 of Lecture Notes in Computer Science, Springer, pp. 103–114.
Handcock, M.S., Raftery, A.E., Tantrum, J.M., 2007. Model-based clustering for social networks (with discussion). Journal of the Royal Statistical Society, Series A 170, 301–354.
Hoff, P.D., Raftery, A.E., Handcock, M.S., 2002. Latent space approaches to social network analysis. Journal of the American Statistical Association 97, 1090–1098.
Hoff, P.D., 2003. Random effects models for network data. In: Breiger, R., Carley, K., Pattison, P. (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, pp. 302–322.
Hoff, P.D., 2005. Bilinear mixed-effects models for dyadic data. Journal of the American Statistical Association 100, 286–295.
Krivitsky, P.N., Handcock, M.S., 2008a. Fitting position latent cluster models for social networks with `latentnet`. Journal of Statistical Software, 24, 2 http://www.jstatsoft.org/v24/i02/.
Krivitsky, P.N., Handcock, M.S., 2008b. `latentnet`: Latent position and cluster models for statistical networks, Version 2.2. Available at: http://statnetproject.org, http://CRAN.R-project.org/package=latentnet.
Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Review 45, 167–256.
de Nooy, W., Mrvar, A., Batagelj, V., 2005. Exploratory Social Network Analysis with Pajek. Cambridge University Press, Cambridge.
Oh, M.S., Raftery, A.E., 2001. Bayesian multidimensional scaling and choice of dimension. Journal of the American Statistical Association 96, 1031–1044.
Oh, M.S., Raftery, A.E., 2007. Model-based clustering with dissimilarities: A Bayesian approach. Journal of Computational and Graphical Statistics 16, 559–585.
Raftery, A.E., Lewis, S.M., 1996. Implementing MCMC. In: Gilks, W.R., Spiegelhalter, D.J., Richardson, S. (Eds.), Markov Chain Monte Carlo in Practice. Chapman and Hall, London, pp. 115–130.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Available at: http://www.R-project.org.
Sampson, S.F., 1969. Crisis in a cloister. Ph.D. Thesis. Cornell University.
Schweinberger, M., Snijders, T.A.B., 2003. Settings in social networks: A measurement model. Sociological Methodology 33, 307–341.
Shortreed, S., Handcock, M.S., Hoff, P.D., 2006. Positional estimation within the latent space model for networks. Methodology 2, 24–33.
Stephens, M., 2000. Dealing with label-switching in mixture models. Journal of the Royal Statistical Society, Series B, Methodological 62, 795–809.
van Duijn, M.A.J., Snijders, T.A.B., Zijlstra, B.H., 2004. p_2: a random effects model with covariates for directed graphs. Statistica Neerlandica 58, 234–254.
White, H.C., Boorman, S.A., Breiger, R.L., 1976. Social structure from multiple networks. I. Blockmodels of roles and positions. American Journal of Sociology 81, 730–780.