

# Positional Estimation Within a Latent Space Model for Networks

Susan Shortreed, Mark S. Handcock, and Peter Hoff

University of Washington, United States

**Abstract.** Recent advances in latent space and related random effects models hold much promise for representing network data. The inherent dependency between ties in a network makes modeling data of this type difficult. In this article we consider a recently developed latent space model that is particularly appropriate for the visualization of networks. We suggest a new estimator of the latent positions and perform two network analyses, comparing four alternative estimators. We demonstrate a method of checking the validity of the positional estimates. These estimators are implemented via a package in the freeware statistical language R. The package allows researchers to efficiently fit the latent space model to data and to visualize the results.

**Keywords:** random graph models, Markov chain Monte Carlo, visualization, maximum likelihood estimation

## Introduction

Social network data is used to study actors and the relationships between them. The concept of social position is canonical in social network theory and has a long history (Faust, 1988; McFarland & Brown, 1973). There are a number of concepts of position in network theory. Lorrain and White (1971) referred to two actors being structurally equivalent if they have identical relations with all actors in a network. Burt (1976) conceptualized a social space where actors who are structurally equivalent occupy the same position in the space. In this context, Burt (1980) defined a social status to be a network position occupied by three or more actors. This conceptualization of position led to the *blockmodels* of White, Boorman, and Breiger (1976) and to the latent class models of Nowicki and Snijders (2001). These related representations of social space have in common the treatment of distance in the space as a measure of dissimilarity of the actors' relational patterns. A distinct conceptualization of social space treats distances as a measure of the degree the actors wish to have relations with one another (Laumann, 1966; McFarland & Brown, 1973). In a 2002 article, Hoff, Raftery, and Handcock (HRH) suggested an approach to modeling networks based on this latter notion of a social space. The HRH model posits the existence of an unobserved latent space of characteristics of the actors and that relationships form as a function of distances between these characteristics. In particular, we focus here on social distances in a Euclidean space.

This article provides an expansion of the practical use of the latent space model by introducing a new estimate for the locations of the actors in social space and comparing it to standard estimates. After reviewing the relevant social network framework and explaining the latent space model and its estimation, we illustrate the model and the new location estimator with two classic social network data sets,

Padgett's Florentine Marriage data (Padgett & Ansell, 1993) and Sampson's Monastery data (Sampson, 1968), in order to investigate parameter estimates.

## Stochastic Models for Social Networks

Social network data typically consist of a set of  $g$  actors and a relational tie  $y_{ij}$ , measured on each ordered pair of actors  $i, j = 1, \dots, g$ . In the most simple cases,  $y_{ij}$  is a dichotomous variable, indicating the presence or absence of some relation of interest, such as friendship, collaboration, or transmission of information or disease. Here we focus on the case of a binary relationship and the  $g \times g$  sociomatrix  $Y = [y_{ij}]$ . We denote  $y_{ij} = 1$  if there exists a relation from actor  $i$  to actor  $j$ , while  $y_{ij} = 0$  will denote that no such directed relation exists. This can be thought of as a graph in which the nodes are actors and the edge set is  $\{(i, j) : y_{ij} = 1\}$ .

The network matrix  $Y$  can be viewed as a random variable with a sample space of  $Y \subseteq \{0, 1\}^n$ , where  $n$  is the total number of possible ties in a network. Each  $Y_{ij}$  can be treated as a Bernoulli random variable with marginal probability  $P(Y_{ij} = 1)$ ; it is this probability, as well as the joint distribution, these methods try to model (Hoff et al., 2002). The family of distributions most commonly used to model social networks is referred to as exponentially parameterized random graphs (Frank & Strauss, 1986). The probability mass function is

$$P_{\eta}(Y = y) = \frac{\exp[\eta^T \mathbf{u}(y)]}{c(\eta)} \quad y \in Y \quad (1)$$

where  $\eta \in \mathfrak{R}^p$  is the model parameter and  $\mathbf{u}: Y \rightarrow \mathfrak{R}^p$  are network statistics (Strauss & Ikeda, 1990). Under this model the  $\mathbf{u}(y)$  are jointly sufficient for  $\eta$  and the model is

the maximum entropy distribution based on the statistics. The normalizing function is

$$c(\eta) = \sum_{y \in Y} \exp(\eta^T \mathbf{u}(y)).$$

A basic example of this model is the Bernoulli random graph. In this density  $u(y_{ij}) = y_{ij}$  and  $\eta_{ij}$  is the logit probability of the  $ij$ th edge occurring. When all of the edges have a common probability, this density simplifies to

$$P(Y = y) = \frac{\exp\left(\eta \sum_{i \neq j} y_{ij}\right)}{c(\eta)} \quad y \in Y$$

where  $\eta = \eta_{ij} = \text{logit}(P(Y_{ij} = 1))$  and  $c(\eta) = (1 + \exp(\eta))^n$ . This can be reduced to  $\log(P(Y = y)) = \eta \mathbf{u}(y) - \log(c(\eta))$ , where  $u(y)$  is the total number of ties in the network.

In many situations  $\eta \in R^p$ , the canonical parameter in Equation 1, is a nonlinear function of a lower-dimensional parameter  $\theta \in R^q$ ,  $q < p$ . In this case we consider a curved exponential family of graph models:

$$P_\theta(Y = y) = \frac{\exp[\eta^T(\theta) \mathbf{u}(y)]}{c(\theta)} \quad y \in Y \quad (2)$$

where usually  $\theta$  is of substantially lower dimension than  $\eta$ . The latent space models we considered in the next section will be considered as curved exponential families, and include the linear exponential models (Equation 1) as a special case.

A complication that arises in social network data is the inherent dependency among ties. For example, if Jon and Peter are friends, and Kyle and Peter are friends, then it is more likely that Kyle and Jon are friends (or that any relationship exists) than if these previous relationships did not exist. This dependency is ignored in the simplest of models—such as the Bernoulli random graph model—and dealt with in various ways in more complex models. Exponentially parameterized models (Equation 1) allow statistics  $\mathbf{u}(y)$  to be chosen to represent complex dependencies (Snijders, Pattison, Robins, & Handcock, in press; Strauss & Ikeda, 1990), and these models can be extended through the use of curved exponential family forms (Hunter & Handcock, in press). Latent space models provide a way to handle dependency via postulating the existence of a *social space*. Various characteristics (usually unknown) dictate actors' positions in the social space, and it is this position that determines their social role and behavior. For example, the concept of a latent space can help model the relationships between Jon, Peter, and Kyle. Suppose that Jon and Peter are friends. This is most likely induced by their closeness in the latent space of characteristics and a higher propensity for friendship between those with similar characteristics (*homophily*). If Peter and Kyle are friends, then it is likely that they too are close to each other in the space of characteristics. The triangle inequality implies Jon and Kyle cannot be too far apart in the space of characteristics. This is, of course, a statistical relationship: closeness in the space does not necessitate friendship, and friendship does not necessitate two individuals being close. If we include distance in this latent space between actors when

modeling the structure of the network, the probability that Jon and Kyle are friends will increase because their relative distance is small. In other words, by modeling the relationships between people as a function of their distance in social space, we are able to take into account potential dependency among relations. While this dependency can also be modeled through a small numbers of statistics and parameters via Equation 1, the latent space model can flexibly capture a wide range of dependence. It is this class of model we describe in the next section.

## The Latent Space Model

The latent space model of HRH is based on two central assumptions. The first states that the probability of a tie between two actors depends on the distance between them in an unobserved social space. The second states that ties between actors occur independently, given their distance apart. Specifically, we assume that each actor has an unobserved position in a  $d$ -dimensional Euclidean latent social space. We then assume that the presence or absence of a tie between two individuals is independent of all other ties, given the positions  $Z = \{z_i\}_{i=1}^g$  in social space of the two individuals:

$$P_\theta(Y = y) = \prod_{i \neq j} P(y_{ij} | z_i, z_j, \beta), \quad (3)$$

where the vector of parameters is  $\theta = (Z, \beta)$ .

We model  $P(y_{ij} | z_i, z_j, \beta)$  using a logistic regression model in which the probability of a tie depends on the Euclidean distance between  $z_i$  and  $z_j$  in social space:

$$\log \text{odds}(y_{ij} = 1 | z_i, z_j, \beta) = \beta - |z_i - z_j|, \quad (4)$$

where  $\log \text{odds}(p) = \log[p/(1 - p)]$ . The distance between actors  $i$  and  $j$  is then  $d_{ij} = |z_i - z_j|$ . The motivation for this method is similar to that for multiple dimensional scaling (MDS) which is widely used as a means of representing the spatial structure of a social network (Breiger, Boorman, & Arabie, 1975). In this context, MDS is a class of methods that can be used to produce a spatial representation of individuals based on similarity or dissimilarity measures between pairs of individuals. Such applications of MDS differ from the model presented here in that MDS is used primarily as a data-analytic means of visualizing given dissimilarities, while our method is a model-based representation of the measured relations and latent locations. The latent space model has the advantage of directly modeling the observed relationship, while the usual choices for dissimilarities in MDS are ad hoc and do not reflect the stochastic nature of the sociomatrix.

Expressed in terms of Equation 2, the model for the graph is

$$P_\theta(Y = y) = \frac{\exp\left(\sum_{i \neq j} \eta_{ij}(\theta) y_{ij}\right)}{\prod_{i \neq j} (1 + \exp(\eta_{ij}(\theta)))}. \quad (5)$$

Hence the parameters are  $\beta$ , the log-odds of a tie for two actors with the same position in social space, and the positions,  $Z$ .

Note that the observations on each directed pair are independent given the positions. Thus we have  $n = g(g - 1)$  independent observations to estimate the  $dg + 1$  parameters (the positions  $Z$  and  $\beta$ ). Note that unconditional on the positions the dyads are dependent. A similar situation occurs in hierarchical or multilevel models.

The true positions are in a dimensionless latent social space with an unknown distance metric. In this article the Euclidean distance metric is used and the actors are positioned in  $\mathbb{R}^d$ . Graphs with different positions for the actors, but which yield the same relative distances, have the same probability under the model. Because distances between nodes are invariant to reflection, rotation, and translation of the  $Z$ 's, there are infinitely many graphs that represent the same relative distances. In order to deal with this, and avoid overestimating the variability in the relative positions of the actors, the Procrustean transformation of a graph is used instead of the graph itself. The Procrustean transformation of a graph  $Z$  around a graph  $Z_0$  changes the graph to have the same reflection, rotation, and translation properties as  $Z_0$ . For more details see Sibson (1979) and Hoff et al. (2002).

## Algorithmic Development

### Theory

In this model we would like to estimate both the positions of the actors in latent space and, in more complex variants, also covariate coefficients. Our proposed estimation technique consists of two approaches. In the first, the parameters are estimated with the maximum likelihood (ML) method by finding the value of the parameters, which maximize the likelihood function (Equation 5). The ML estimation procedure finds parameter values that give the highest likelihood to the data as observed.

The second approach is a Bayesian analysis with a diffuse prior ( $\pi$ ) for  $\theta$ . In many situations where the latent space model is used, only vague information is known about the parameters. Here the priors are independent normal distributions centered around zero, with a large variance. In this situation the data should drive the posterior with the prior distribution having little effect. For more information on Bayesian statistical methods, see Gelman, Carlin, Stern, and Rubin (1995). To estimate the posterior distribution of the parameters we implement a Markov chain Monte Carlo (MCMC) algorithm. Given certain conditions (all met here), the MCMC will converge to the posterior distribution, which means that the chain will be taking samples from the joint posterior distribution of the parameters. These samples can then be used to make inferences about the parameter values and their marginal posterior distributions. The general algorithm used is as follows (Hoff et al., 2002):

1. Find the maximum likelihood estimate (MLE) for  $\theta = (\beta, Z)$ .
2. Run an MCMC chain.

To start, set  $k = 0$  and  $Z^0$  to be the MLE of the  $Z$  matrix.

Repeat steps (a)–(d)  $N \times M$  times, storing  $\{Z^N, \beta^N\}_{j=1}^M$ . Here  $N$  is the interval between samples and the result is a sample of size  $M$  from the posterior:

(a) Sample a proposal  $\tilde{Z}$  from independent multivariate Gaussians centered around  $Z^k$ .

(b) With probability equal to the minimum of 1 and  $\frac{P(Y|\tilde{Z}, \beta^k)\pi(\tilde{Z})}{P(Y|Z^k, \beta^k)\pi(Z^k)}$

- accept  $\tilde{Z}$  to be  $Z^{k+1}$ .
- otherwise  $Z^{k+1} = Z^k$ .

(c) Sample a proposal  $\tilde{\beta}$  from a Gaussian distribution centered at 0.

(d) With probability equal to the minimum of 1 and  $\frac{P(Y|Z^k, \tilde{\beta})\pi(\tilde{\beta})}{P(Y|Z^k, \beta^k)\pi(\beta^k)}$ ,

- accept to be  $\beta^{k+1}$ .
- otherwise  $\beta^{k+1} = \beta^k$ .

3. For each of the  $M$  samples of the posterior position matrix ( $\{Z^N\}_{j=1}^M$ ), store its Procrustean transformation relative to the  $Z_{mle}$ . Denote these by  $\{\tilde{Z}^N\}_{j=1}^M$  where the transformation  $\tilde{Z}$  of  $Z$  relative to  $Z_{mle}$  is  $Z_{mle}Z^T(Z_{mle}^T Z_{mle} Z^T)^{-1/2}Z$ .

Note that steps 2(a) and (b) comprise a standard Metropolis-Hastings step to update the positions, and steps 2(c) and (d) are a standard Metropolis-Hastings step to update the  $\beta$  (Robert & Casella, 2005). This algorithm is similar to those used for other latent variable models (Skrondal & Rabe-Hesketh, 2004) and is modified from the algorithm implemented by HRH, where the Procrustean transformation of the position matrix was taken during the chain after acceptance.

### Four Alternatives for Point Estimation of the Positions in Latent Space

In many situations point estimates for  $\beta$  and the positions are desired in addition to their posterior distributions. There are three natural candidates for point estimates for the parameters in the model: the MLE, the posterior mean, and the posterior mode. However, as the model specifies the distances between actors only and not their locations, these estimates may be very poor representations of the positions. That is, even when we expect the distance between actors to be accurately determined, the locations themselves are not. As we shall see in the next section, the above three estimators can give poor estimates of the locations. To resolve this, we propose a fourth estimator, the values of  $(Z, \beta)$  that minimize the Kullback-Leibler divergence to the mean posterior distance model. This will be referred to as the MKL estimate.

The Kullback-Leibler divergence of a distribution with probability mass function  $p$  from the distribution with probability mass function  $q$  is

$$E_q[\log(q) - \log(p)].$$

Let  $\phi$  and  $\eta$  be alternative latent space parameters for Equation 5. The Kullback-Leibler divergence,  $KL(\phi, \eta)$ , of

the latent space model with parameter  $\eta$  from the parameter  $\phi$  is

$$\begin{aligned} & E_{Y|\phi} \left[ \log \left( \frac{P_{Y|\phi}(Y = y)}{P_{\eta}(Y = y)} \right) \right] \\ &= \sum_{y \in Y} \log \left( \frac{P_{Y|\phi}(Y = y)}{P_{\eta}(Y = y)} \right) P_{Y|\phi}(Y = y) \\ &= \sum_{y \in Y} (\phi - \eta)^T y P_{Y|\phi}(Y = y) + \log \left( \frac{c(\eta)}{c(\phi)} \right) \\ &= (\phi - \eta)^T E_{Y|\phi}[Y] + \log \left( \frac{c(\eta)}{c(\phi)} \right). \end{aligned}$$

The MKL estimates are the parameter values,  $\eta$ , minimizing the posterior mean of the Kullback-Leibler divergence:

$$\begin{aligned} & E_{\phi|Y_{obs}}[KL(\phi, \eta)] \\ &= \phi^T E_{\phi|Y_{obs}}[E_{Y|\phi}[Y]] - \eta^T E_{\phi|Y_{obs}}[E_{Y|\phi}[Y]] \\ &\quad - E_{\phi|Y_{obs}}[\log(c(\phi))] + E_{\phi|Y_{obs}}[\log(c(\eta))] \\ &= \phi^T E_{\phi|Y_{obs}}[E_{Y|\phi}[Y]] - \eta^T E_{\phi|Y_{obs}}[E_{Y|\phi}[Y]] \\ &\quad - \log(c(\phi)) + \log(c(\eta)). \end{aligned}$$

Since  $E_{\phi|Y_{obs}}[E_{Y|\phi}[Y]] = E[Y|Y_{obs}]$  is the mean posterior probability of a tie under the model with parameter  $\phi$  and the first and third terms do not depend on  $\eta$ , the optimization problem is simplified to finding the parameter values that maximize

$$\frac{\exp(\eta^T E[Y|Y_{obs}])}{c(\eta)}$$

This optimization is easily implemented using the likelihood routines already used in the algorithm. The posterior mean  $E[Y|Y_{obs}]$  can be accurately estimated from the MCMC samples and does not require the positions, but only the distances.

For all four estimators of the positions and  $\beta$  the estimates are free to move around the parameter space without restriction. Thus, the positional estimates are similar but are stretched or contracted in comparison to one another. This lack of restrictions on the parameter space when maximizing the likelihood usually results in the nodal positions being pushed far away from one another with  $\beta$  increasing to compensate for the large distances. The prior distributions in the MKL method force more information to be present from the data to drastically increase  $\beta$ ; thus the absolute distances are kept reasonably small. In order to compare these estimates, each position was **normalized** by setting the coordinates equal to  $Z_{ij}^{norm} = Z_{ij} / \sqrt{\sum_{j=1}^k Z_{ij}^2}$

## Application of the Latent Space Model to Two Networks

This section describes the analysis of two data sets using the latent space model and the method described above. In each of the analyses, several parameter values for the proposal distributions were tried, and the values that appeared to allow the parameters to move around the space without compromising the acceptance rates too greatly were chosen. Each data set is fitted with the latent space model, and the parameter estimates are investigated.

### Florentine Marriage Data

The first data set discussed is known as the Florentine Marriage data. This data was compiled by Padgett and Ansell (1993), and focuses on 16 prominent Florentine families and their business and marriage relations. The data collected by Padgett covers a large time span as well as types of networks, but this analysis will concentrate on the marriage data from the fifteenth century. Each actor in the data set is a family, and a tie is present if there existed at least one marriage between the families. For easier visualization the isolates were removed from the data before analysis; thus this analysis will be on 15 of the original 16 families. The Markov chain takes a sample of size 700 from the posterior distribution of the parameters, allowing 10,000 iterations for burn-in and 1,000 iterations between samples. The standard deviation parameters for the proposal distributions of the chain were  $\delta_\beta = 0.5$  and  $\delta_z = 0.3$ . These values only affect the efficiency of the sampler by altering the deviation from proposal to proposal of the parameters and were chosen by monitoring the trace plots of the MCMC for mixing. The prior distribution on  $\beta$  is  $N(0, 10^2)$  and the prior distribution on each of the coordinates  $Z_{ij}$  is also  $N(0, 10^2)$ . The plots of the log-likelihood values and  $\beta$  at each sampled interval in the chain, seen in Figure 1, indicate convergence. The bottom two plots of Figure 1 are the marginal posterior distribution of  $\beta$  and the positions of the actors.

Figure 2 displays the four different normalized estimates for the  $\mathbb{R}^2$  positions. The basic structure of the graph is similar in all four estimates; the main difference is located on the left-hand side of the graphs: actors 1, 10, and 13. These actors do not have many ties to other actors, and so the model expects that they are distant from the other actors. However, this does not tell us where they are exactly, only that they are not likely to have positions close to the positions of the other actors. Thus, we know their distances are large, but there is limited information contained in the network about their position in latent space. While the posterior distances are stable, the positions are not; the posterior samples sometimes place actor 1 on top and actor 13 on the bottom, sometimes vice versa. This results in a bimodal posterior distribution for the positions of these actors. In Figure 1, it can be seen that the (green) points corresponding to actor 1, the (pink) points corresponding to actor 13, and the (blue) points corresponding to actor 10 form two distinct clusters each. (Color versions of Figure 1 and Figure 4 are available at <http://www.csss.washington.edu/Published>.) It is well known that the mean and marginal modes of joint distributions are poor point estimates for central tendencies of multimodal distributions. This multimodality implies the MLE and MKL estimates better summarize the estimated positions of the actors in this example. While the normalized graphs are comparable, the raw graphs are quite different: The square root of the sum of squared distances in the MLE graphs is 68.63 and the MLE for  $\beta$  is 14.45. The same quantities for the MKL method are 20.94 and 3.96. This dramatic difference between estimates demonstrates that the interpretation of  $\beta$  is meaningless without the distances.



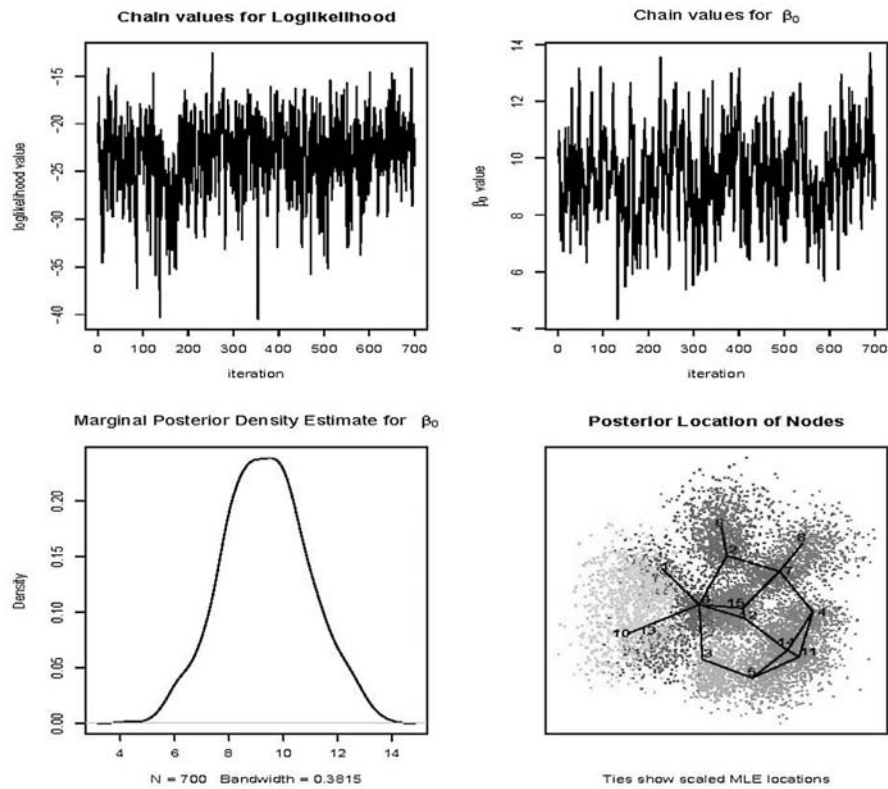


Figure 1. Labeled clockwise starting from upper left: (a) Chain values for the log-likelihood. (b) Chain values for  $\beta$ . (c) Marginal Posterior for  $\beta$ . (d) Marginal posterior for positions.

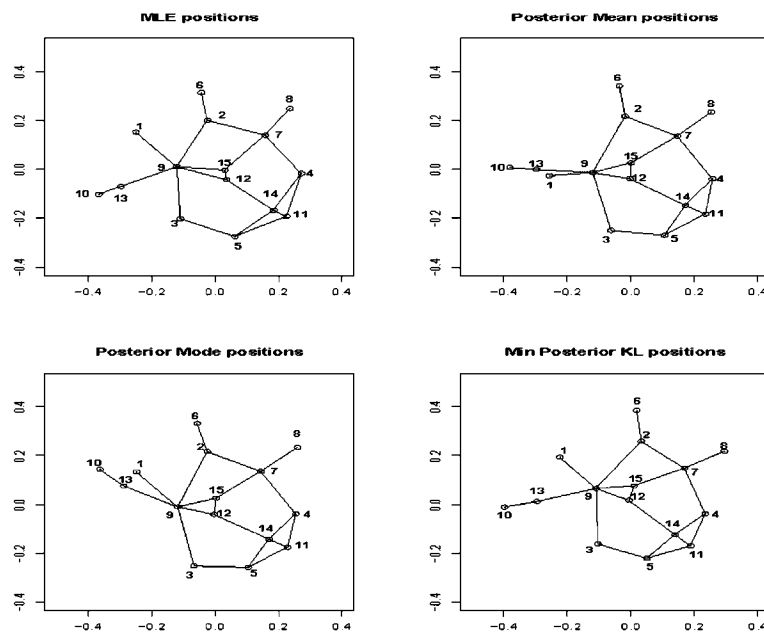


Figure 2. Positional estimates for the Florentine data. While the basic structure is similar, it appears that there is some discrepancy in the location of actors 1, 10, and 12.

The number of ties in a graph is a basic network summary statistic, and a good parameter estimate should produce graphs in which the observed number of ties (20 undirected ties in the Florentine data) has a high probability. Figure 3 shows the density estimates of the number of ties produced from a probability distribution corresponding to the latent space model with each of the four parameter estimates. The MLE and MKL estimates appear to produce reasonable distributions on the number of ties, while the posterior mean and mode tend to generate graphs that have “too many” ties.

The better performance of the MKL could be due to the bimodality of the posterior distribution of some actors' positions. In Figure 2, it is noted that actors 1, 10, and 13 are closer in the posterior mean and mode representation than in the MLE and MKL representation; thus they have a smaller distance when these first two positional estimates are used. Focusing attention on the troublesome actors (1, 10, 13) and the ties between them shows some differences in the graphs generated by each of the estimates found in Figure 6. The MLE estimates produce graphs that have on average 1.25 ties between the three actors, the posterior mean graphs have on average 2.98 ties, the posterior mode graphs have on average 2.98 ties, and finally, the MKL graphs have on average 1.48 ties. In the observed graph there is one tie between these three actors (1–13), so the MLE and MKL estimates produce graphs with ties between this trio most like the observed graph. For this analysis it appears that the MLE and MKL estimates are the best point estimates for the coefficients and positions of the actors.

## Monk Data

The second analysis discussed is of another standard data set in the social network literature. In 1968, Sampson collected data on 18 monks and their interpersonal relations. Each monk was asked about positive relations with the other monks, and reciprocity was not required; thus the graph is directed. The data contain 56 directed ties between the 18 monks. Similar to the Florentine analysis, the chain samples 700 points from the posterior, allowing 10,000 iterations for burn-in and 1,000 iterations between samples. In the monk analysis the standard deviation parameters for the proposal distributions in the chain were  $\delta_\beta = 0.3$  and  $\delta_z = 0.2$ . The prior distribution on  $\beta$  was  $N(0, 5^2)$  and the prior distribution on the positions coordinates,  $Z_{ij}$ , is  $N(0, 10^2)$ . Figure 4 contains the summary plots of the Markov chain. The chain values for the log-likelihood and  $\beta$  indicate convergence. The marginal density of  $\beta$  appears not to be skewed and to be unimodal. The marginal posterior for the positions appears to show clustering of the position values, although some of the clusters appear to be elongated, suggesting uncertainty in the locations.

The four types of positional estimates are given in Figure 5 with the same normalization technique used earlier for easier comparison. The differences between these estimates are more complex than in the Florentine analysis. Some of the actors seem to have stable positions regardless of the method used: actors 1, 2, 4, 5, 6, 9, 11, and 18. Actors 8 and 12 seem to have some uncertainty that may result in a bimodal distribution since they have collapsed

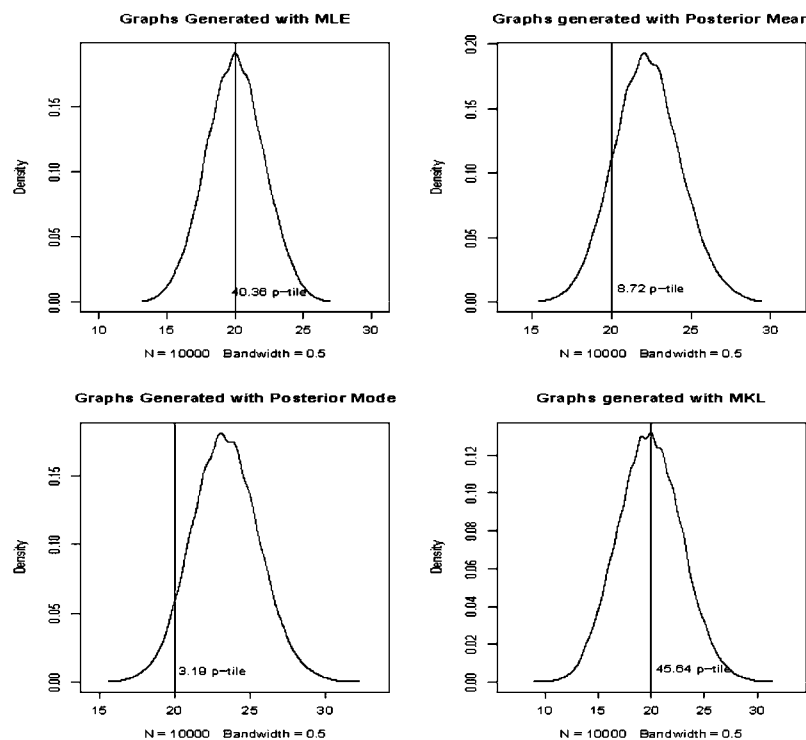


Figure 3. Density estimates for the number of ties in a graph generated by the latent space model with the specified estimates as parameters.

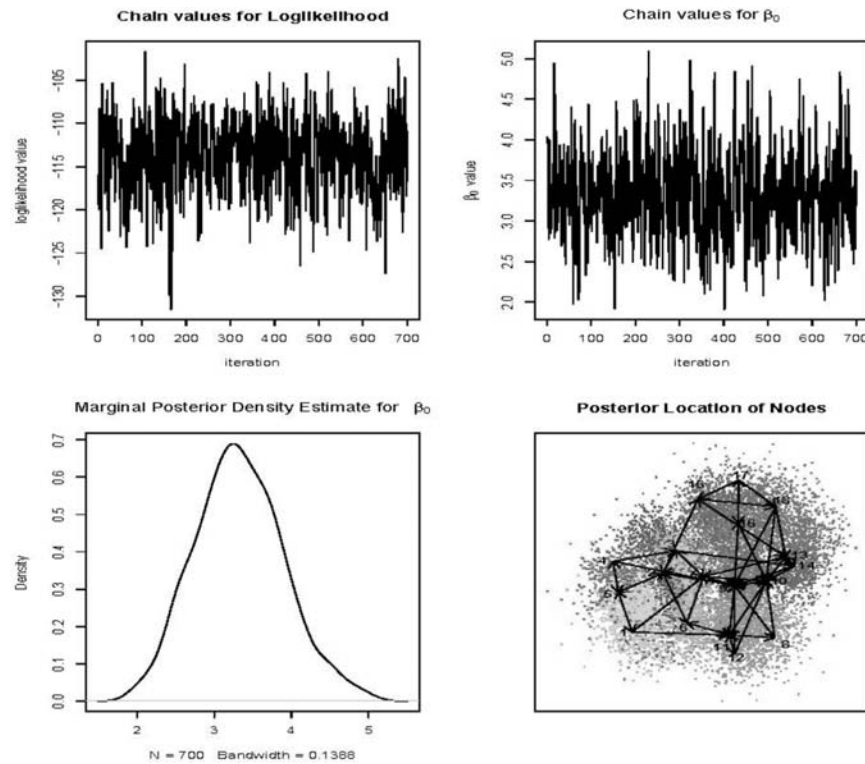


Figure 4. Labeled clockwise starting from upper left: (a) Chain values for the log-likelihood. (b) Chain values for  $\beta$ , (c) Marginal posterior for  $\beta$ , (d) Marginal posterior for positions.

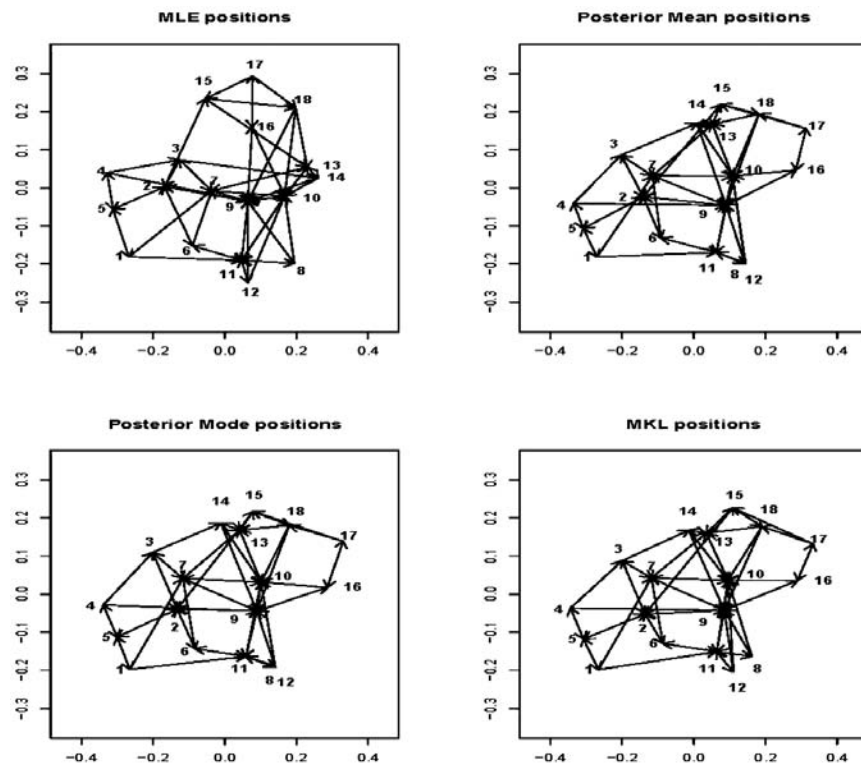


Figure 5. Positional estimates for monk data. While there is some similarity in the general structure of the graphs, several actors are given quite different positions under the different estimates.

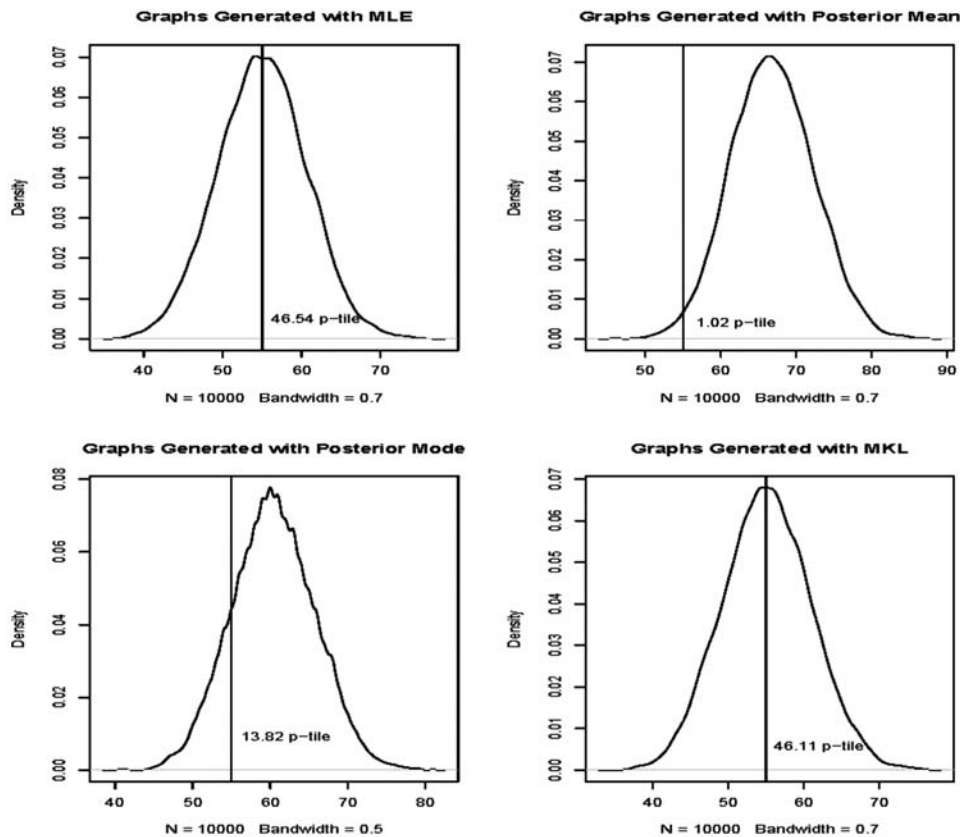


Figure 6. Density estimates for the number of ties in a graph generated by the latent space model with the specified estimates as parameters.

onto one another in the posterior mean and mode estimates, yet the MLE and MKL keep them close yet distinct. Actor 3 is closer to the center of the graph in the MLE positions, but in the other three it has moved up and taken actor 7 with it. Actors 13 and 14 also appear to be a pair that move together; in the MLE estimate they are at the right side of the graph in the center height-wise, yet in the other posterior estimates they are both up near the top of the graph. It appears that actor 10 travels with these actors into the center of the graph. Actor 17's position also shifts from the center top of the graph in the MLE to the upper-right-hand corner. These plots as well as the marginal posterior position plots seem to imply that the positional distribution is not very well behaved. Once again the MLE and MKL estimates would appear to be better estimates than the posterior mean and mode because of this complex behavior. This analysis produced MLE and MKL estimates that are more similar; the MLE estimate for  $\beta$  is 2.38, with a graph with the square root of sum of squared distances of 15.77, while the MKL estimated the same quantities with 1.62 and 12.54.

The superiority of the MLE and MKL can be seen once again in these density plots for the number of ties in graphs generated by each of the estimates found in Figure 6. The MLE and MKL seem to generate graphs that tend to have close to the observed value of 56 ties, while the posterior mean and mode estimates generate graphs with many more ties.

## Discussion

This article has provided a brief introduction to the latent space model for modeling network data as well as code that allows users to fit these models in the statistical free-ware R. The code provided in the package *latentnet*<sup>1</sup> is slightly different than the model originally presented by HRH in that all coefficients have a Gaussian prior, including the intercept. This code also expands on the point estimates for the coefficients and positions discussed in HRH, by providing all four of the estimators described: MLE, posterior mean, posterior mode, and the estimator that minimizes Kullback-Leibler divergence from the posterior.

1 The *latentnet* package is written in a combination of (the open-source statistical language) R and (ANSI standard) C, and runs on Linux, UNIX, Windows, and Macintosh environments. It is downloadable from the Internet (R Development Core Team, 2005). Details on installation of the *latentnet* library (and the network library that it requires) can be found in Handcock, Hunter, Butts, Goodreau, and Morris (2003).



Of the four explored here, the estimates that seem to produce graphs most consistent with the observed network are the MLEs and the MKL divergence estimates. It was shown that by definition the MKL is designed to minimize the posterior mean of the Kullback-Leibler divergence from the true model. The MLE of the distances can be shown to minimize the Kullback-Leibler divergence from the data (Barndor-Nielsen, 1978). It follows that the MLE of the positions are the values that minimized the Kullback-Leibler divergence from the MLE distances. That is,

$$\eta_{MLE} = \operatorname{argmax}_{\eta} \{P_{\eta}(Y = y)\} = \operatorname{argmin}_{\eta} \{KL(\hat{\phi}, \eta)\},$$

where  $\hat{\phi}$  is the MLE of the data. The MKL minimizes the Kullback-Leibler divergence from the model with parameters given by the posterior expectation of the graph under the mean-value parameterization of the exponential family model (Handcock, 2003). Similarly, the MLE minimizes the Kullback-Leibler divergence from the model with mean-value parameter given by the observed graph. In this sense, the MKL is the closest model to the posterior expectation of the graph, while the MLE is the closest to the observed graph. Heuristically, if the model is approximately correct, one expects that the posterior mean graph will be closer to the true model than the observed graph due to statistical averaging and use of prior information. In practice, the MKL estimate will usually be superior to the MLE, as the averaging reduces the first-order statistical variation while inducing a second-order bias due to misspecification. However, both should be considered, as they represent contrasting assumptions about the model (Handcock, 2003).

In network data, the purpose is to model the joint probability of all the ties in the network as well as the probability of a tie between any two actors. While a graph has  $g$  actors and  $g(g-1)$  directed ties, it is only one realization of the whole network. This is similar to time series data where we only have one sample of the whole series but many dependent observations within the series. In many such statistical situations, the behavior of the MLE is understood under both large and finite sample sizes. These illustrate that our intuition about sample size is altered when the data is dependent. Working with a small network forces more emphasis to be placed on the model than the observations. In both Padgett's Florentine and Sampson's monk data the MLE and MKL estimates appear to behave similarly. As the size of the graph grows and ties are sparse, it is reasonable to believe that the unrestrictedness of the MLE could lead to instability in the estimates. While the MKL method may introduce bias into the estimates by putting more trust in the model, it may also produce estimates with less variability.

This article proposes a new estimator and examines some of its properties when applied to two data sets. More research is needed to truly understand the behavior of each of the point estimates (especially the variance) as well as to develop meaningful error bounds. It is unclear how the estimators will perform as the number of actors grows, particularly since the dimension of the parameter space grows with each additional actor. In order to understand the loss of flexibility and the gain in modeling ability of using the

MKL estimates, it is necessary to understand the impact of the prior on the posterior distribution of the parameters.

## Acknowledgments

The authors are grateful to Martina Morris, Adrian E. Raftery, Marijtje van Duijn, and Jeroen Vermunt for numerous helpful suggestions. This research is supported by Grant DA012831 from NIDA and Grant HD041877 from NICHD.

## References

- Barndor-Nielsen, O. E. (1978). *Information and exponential families in statistical theory*. New York: Wiley.
- Breiger, R. L., Boorman, S. A., & Arabie, P. (1975). An algorithm for clustering relational data, with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12, 328–383.
- Burt, R. S. (1976). Position in networks. *Social Forces*, 55, 93–122.
- Burt, R. S. (1980). Models of network structure. *Annual Review of Sociology*, 6, 79–141.
- Faust, K. (1988). Comparison of methods for positional analysis: Structural and general equivalence. *Social Networks*, 10, 313–341.
- Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81, 832–842.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall.
- Handcock, M. S. (2003). *Assessing degeneracy in statistical models of social networks* (Working Paper No. 39). Seattle, University of Washington, Center for Statistics and the Social Sciences.
- Handcock, M. S., Hunter, D., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). Statnet: An R package for the statistical modeling of social networks. Retrieved from <http://www.csde.washington.edu/statnet>.
- 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.
- Hunter, D. R., & Handcock, M. S. (in press). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*.
- Laumann, E. O. (1966). *Prestige and association in an urban community: An analysis of an urban stratification system*. New York: Bobbs-Merrill.
- Lorrain, F., & White, H. (1971). Structural equivalence of individuals in social networks blockstructures with covariates. *Journal of Mathematical Sociology*, 1, 49–80.
- McFarland, D. D., & Brown, D. J. (1973). Social distance as a metric: A systematic introduction to smallest space analysis. In E. O. Laumann (Ed.), *Bonds of pluralism: The form and substance of urban social networks* (pp. 213–253). New York: Wiley.
- Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96, 1077–1087.
- Padgett, J. F., & Ansell, C. K. (1993). Robust action and the rise of the Medici. *American Journal of Sociology*, 98, 1259–1319.
- R Development Core Team. A language and environment for statistical computing [Computer software]. Vienna, Austria:

- R Foundation for Statistical Computing. Retrieved from <http://www.csde.washington.edu/statnet>.
- Robert, C. P., & Casella, G. (2005). *Monte Carlo statistical methods* (2nd ed.). New York: Springer-Verlag.
- Sampson, S. F. (1968). *A novitiate in a period of change: An experimental and case study of relationships*. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.
- Sibson, R. (1979). Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the Royal Statistical Society, Series B, Methodological*, 41, 217–229.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman and Hall.
- Snijders, T. A. B., Pattison, P., Robins, G. L., & Handcock, M. S. (in press). New specifications for exponential random graph models. *Sociological Methodology*.
- Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85, 204–212.
- 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.

---

Susan Shortreed

---

Department of Statistics  
University of Washington  
Box 354322  
98195-4322 Seattle, WA  
USA  
Tel. +1 206 543 7237  
Fax +1 206 685 7419  
E-mail [susanms@stat.washington.edu](mailto:susanms@stat.washington.edu)

---