Journal of Educational and Behavioral Statistics
2013, Vol. 38, No. 3, pp. 295–318
DOI: 10.3102/1076998612458702

Hierarchical Network Models for Education Research: Hierarchical Latent Space Models

Tracy M. Sweet
Andrew C. Thomas
Brian W. Junker
Department of Statistics, Carnegie Mellon University

Intervention studies in school systems are sometimes aimed not at changing curriculum or classroom technique, but rather at changing the way that teachers, teaching coaches, and administrators in schools work with one another—in short, changing the professional social networks of educators. Current methods of social network analysis are ill-suited to modeling the multiple partially exchangeable networks that arise in randomized field trials and observational studies in which multiple classrooms, schools, or districts are involved, and to detecting the effect of an intervention on the social network itself. To address these needs, we introduce a new modeling framework, the Hierarchical Network Models (HNM) *framework. The HNM framework can be used to extend single-network statistical network models to multiple networks, using a hierarchical modeling approach. We show how to generalize the latent space model for a single network to the HNM/multiple-network setting, and illustrate our approach with real and simulated social network data among education professionals.*

Keywords: *social network analysis, professional communities, field trials in education research, Bayesian modeling, partial exchangeability*

1 Introduction

Social network analysis (SNA) is a collection of quantitative methods for comparing and measuring relationships among individuals in a network. Goldenberg, Zheng, Fienberg, and Airoldi (2009) trace the development of modern social network analysis since 1959, though social networks have been the object of formal study as far back as Moreno (1934). Individuals and their relationships can be represented visually as directed or undirected graphs; the nodes represent individuals, and the edges represent asymmetric or symmetric relationships—or ties—between pairs of individuals. Two such networks, depicting asymmetric advice-seeking behavior regarding two different subjects among a group of teachers, are depicted in Figure 1.

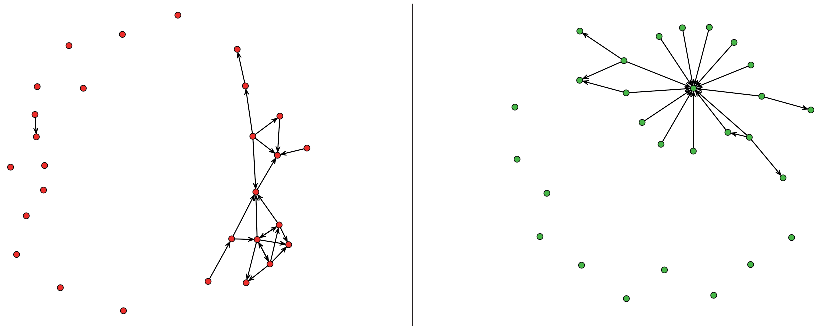

FIGURE 1. *Two social networks, depicting asymmetric advice-seeking behavior among groups of teachers regarding reading (left) and math (right), from Pitts and Spillance (2009). Vertices, or nodes, represent individual teachers. Arrows, or directed edges, point from advice seeking teachers to advice providing teachers.*

While social networks have been used to describe the structure of data in education research, they are seldom analyzed with fully specified statistical models. Instead, social network structure is often analyzed qualitatively to assess its influence on other factors of interest (e.g., Lin, 1999) or summarized to produce measures of social capital (Frank, Zhao, & Borman, 2004; Penuel, Riel, Krause, & Frank, 2009) and other inputs for further quantitative and qualitative analyses (e.g., Bonsignore, Hansen, Galyardt, Aleahmad, & Hargadon, 2011; Coburn & Russell, 2008; Thomas, 2000). Although both student (Harris et al., 2009) and teacher (Pitts & Spillane, 2009) social networks are of interest in education research, we focus on professional social networks among teachers.

Professional social networks in schools highlight the working relationships among staff members and can provide insight into the mechanisms that affect instructional quality and student outcomes. In addition, social network structure may be an important intermediate variable in education intervention research. Many interventions (e.g., Glennan & Resnick, 2004; Hord, Roussin, & Sommers, 2010; Matsumura, Garnier, & Resnick, 2010; McLaughlin & Talbert, 2006; Spillane, Correnti, & Junker, 2009) are applied to whole schools, or grades within schools, and target changes in professional social structure such as increasing teacher collaboration, creating or maintaining small learning communities, or changing the distribution or structure of leadership.

Social structure can provide insight into how an intervention is taken up in a school system (e.g., Matsumura et al., 2010). In addition, particular forms of social network structure are central to *theories of action* for enhancing instructional quality directly (Hord et al., 2010) or as part of a larger organizational framework for effective schools and districts (Glennan & Resnick, 2004). Existing research on professional social networks and educational interventions supports these claims even though these studies tend to focus on descriptive statistics (as catalogued by Kolaczyk, 2009, chap. 4).

For example, Moolenaar, Daly, and Sleegers (2010) find an association between teacher willingness to invest in change and the centrality of their principal, where centrality is defined as the number of teachers seeking advice from their school's principal. Daly, Moolenaar, Bolivar, and Burke (2010) study the number of connections among subgroups of teachers who taught the same grade, finding that dense subgroups were involved with the intervention at a greater depth than sparse subgroups. In fact, Frank, Zhao, and Borman (2004) and Penuel, Frank, and Krause (2006) both find that access to expertise through the network was correlated with change in teacher practice. Similarly, Frank, Penuel, Sun, Chong, and Singleton (2010) find that teachers were more likely to employ a particular teaching practice if they were connected to teachers who also employed that practice. Conversely, Weinbaum, Cole, Weiss, and Supovitz (2008) find, in an observational study of social networks among teachers in 15 schools, that schools that were involved in some type of schoolwide initiative had more ties among teachers. Penuel et al. (2010) study two schools involved in a schoolwide initiative at two time points, and document change in network structure in each school, possibly associated with the initiatives.

As these examples suggest, most such studies utilize descriptive statistics of network measures, either as quantities of direct interest or as covariates in a linear model. For example, Daly et al. (2010) compare teachers within each grade across schools by reporting the mean number of in-ties (i.e., the mean number of times each teacher is identified by a peer as a connection), or mean in-degree, for each of the five schools studied. Frank et al. (2004) and Penuel et al. (2006) use a linear combination of the number of out-ties for each teacher as a covariate in an ordinary least squares (OLS) regression model, pooling teachers across schools. Similarly, Moolenaar et al. (2010) use the number of in-ties for each principal as a predictor in a hierarchical linear model (HLM), nesting teachers within schools. Frank et al. (2010) first use friendship network ties to cluster teachers into subgroups and then use a linear combination of the number of out-ties as a predictor in a three-level HLM, nesting teachers within subgroups within schools.

Although descriptive statistics of the network can be informative, using them alone risks missing important features of the full network. For example, if the size of the network (e.g., number of teachers) is not constant, the number of ties must be interpreted in terms of the size of the network. This issue is more pronounced when working with multiple networks, for example in an intervention study involving several treatment and control schools of differing sizes. Even when the networks have approximately the same size, as in Figure 1, the topology of ties may be quite different. For example, a tie from an advice seeker to a popular advisor may be structurally different than a tie to a more isolated teacher. It is not at all obvious how to keep such information in the system even by considering additional descriptive statistics, such as the proportion of teachers without ties. Moreover, dependence between ties within a network will affect comparisons of descriptive statistics between networks.

An alternative approach involves building a statistical social network model, formalizing the likelihood of observing a particular network from the space of all possible networks. Among statistical network models currently receiving intensive interest in the literature, we consider exponential random graph models (ERGMs), latent space models (LSMs), and mixed membership stochastic block models (MMSBMs). These will be described in more detail in Section 2.1.

Although software exists to fit each of these network models to real data, they have seldom been used in education research. Exceptions include two policy studies, Penuel et al. (2010) and Weinbaum et al. (2008), who fit p_2 models (Lazega & van Duijn, 1997), a type of exponential random graph models (ERGM), to study communication among teachers regarding an initiative. Both studies analyze types of communication across several schools by fitting a separate, single-network model for each school. But the models for different schools are not linked in any way, and comparisons are based exclusively on qualitative assessments of the parameter estimates in each model.

There is a good reason for this: Existing social network models are mostly inadequate for the types of problems studied in education research. Comparing different treatment conditions in an intervention study requires models that can accommodate at least two networks, as well as parameters for treatment effects. Moreover, education interventions generally involve several schools (i.e., professional social networks) in each condition, but again existing social network models are largely confined to fitting one network at a time.

To address these needs, we introduce a new modeling framework, the *Hierarchical Network Models* (HNM) framework. This framework allows us to borrow strength across multiple partially exchangeable networks for parameter estimation, as well as pools information from multiple networks to assess treatment and covariate effects. Other features modeled in single network models, for example, reciprocity and transitivity, can also be incorporated into these models but we will not address features. Some pioneering work with multilevel structures for multiple networks has been done by Zijlstra, van Duijn, and Snijders (2006), who introduce a multilevel p_2 model; Snijders and Kenny (1999), who extend the social relations model to multiple levels; and Templin, Ho, Anderson, and Wasserman (2003), who propose a multilevel model for ERGMs, but our proposed framework is more general. HNMs can accommodate any ERGM, LSM, MMSBM, or any other single-network model in a multiple-network setting, as well as essentially arbitrary network level experimental interventions.

In Section 2, we briefly review the three modeling families mentioned earlier and introduce the general HNM framework. Although this framework is quite general, we focus for specificity on hierarchical latent space models (HLSMs) in Section 3. We also briefly consider similarities and differences between

HNMs and HLMs (Raudenbush & Bryk, 2002). In Section 4, we explore fits of HLSMs to real and simulated professional social network data in education. Finally, in Section 5, we indicate some future directions.

2 Modeling Framework

2.1 Single-Network Models

A single social network Y among n individuals or actors can be represented by an n × n *tie matrix*

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nn} \end{bmatrix}, \tag{1}$$

where Y_{ij}, $i, j \in \{1, \ldots, n\}$, is the value of the tie or edge from actor i to actor j. In a network graph, such as those in Figure 1, each individual i is represented by a node, and each nonzero Y_{ij} is represented by an edge between nodes i and j. If the relationship is symmetric, like teacher collaboration, the ties are shown as simple line segments and Y is a symmetric matrix. If the relationship is asymmetric, like advice seeking, the ties are drawn as arrows or directed edges, as in Figure 1, and the matrix Y need not be symmetric.

If we only collect data on the absence or presence of a tie, then

$$Y_{ij} = \begin{cases} 1, & \text{if there is a tie from } i \text{ to } j \\ 0, & \text{else} \end{cases}.$$

If instead Y_{ij} is a measure of tie strength, such as a count value or correlation, then Y_{ij} may be an integer or real number on the positive real line, $Y_{ij} \in [0, \infty)$. Other possibilities also exist,[1] but in this article we restrict ourselves to binary or real-valued Y_{ij}.

ERGMs (Wasserman & Pattison, 1996) represent the probability of a social network as a function of existing network statistics and individual and tie covariates. A general ERGM may be expressed as

$$P(Y|\theta) = \frac{1}{C(\theta)} \exp\left\{ \sum_{p=1}^{P} \theta_p^T s_p(Y) \right\}, \tag{2}$$

where $(s_1(Y), \ldots, s_P(Y))$ is a set of P sufficient statistics, typically descriptive network statistics such as number of ties and other network structures; $(\theta_1, \ldots, \theta_P)$ are the corresponding parameters to be estimated for each statistic; and $C(\theta)$ is a normalizing constant.

LSMs (Hoff, Raftery, & Handcock, 2002) assume each individual in the network occupies a position in a latent social space, analogous to multidimensional scaling or latent proximity preference models. The model represents the likelihood of a tie as a function of covariates and latent space positions, for example, teachers who are far apart in this social space are less likely to have a tie between them. A general LSM for Bernoulli ties may be expressed as

$$\operatorname{logit} P[Y_{ij} = 1] = \beta^T \mathbf{X}_{ij} - d(Z_i, Z_j), \tag{3}$$

where $\mathbf{X}_{ij}$ is a vector of covariates relating to Y_{ij}, β is a set of regression coefficients to be estimated, Z_i and Z_j are latent location variables for individuals i and j, and $d(z_1, z_2)$ is a distance function; for example, $d(z_1, z_2) = |z_1 - z_2|$. The Z_is are interpreted as locations of the nodes in a low-dimensional Euclidean latent space, analogous to the latent locations in a multidimensional scaling or a latent proximity preference model (Johnson & Junker, 2003; Polak, Heiser, De Rooij, & Busing, 2003). These positions contribute to the model through their pairwise distances, which are invariant to rotation and translation. The probability of a tie increases as these locations become closer together, and may also be affected by observable covariates $\mathbf{X}_{ij}$. Clearly, Equation 3 is one member of a class of mixed generalized linear models (GLMs) that can be adapted to other discrete or continuous tie values Y_{ij} by changing the link function and error distribution.

Finally, MMSBMs (Airoldi, Blei, Fienberg, & Xing, 2008) allow individuals to move between latent classes or blocks, as ties with other individuals are considered; the likelihood of a tie between two individuals is determined by their group membership. A general MMSBM for Bernoulli tie data may be expressed as

$$\begin{aligned} Y_{ij} &\sim \text{Bernoulli}\,(S_{i\to j}^T B R_{j\leftarrow i}) \\ S_{i\to j} &\sim \text{Multinomial}\,(1, \boldsymbol{\theta}_i) \\ R_{j\leftarrow i} &\sim \text{Multinomial}\,(1, \boldsymbol{\theta}_j), \end{aligned} \tag{4}$$

where the vectors $\boldsymbol{\theta}_i$ of multinomial probabilities and the matrix B of Bernoulli probabilities are to be estimated. $S_{i\to j}$ indicates which latent class or block node i belongs to, when *sending* a tie to j. Similarly, $R_{j\leftarrow i}$ indicates which latent class (block) node j belongs to, when *receiving* a tie from i. Thus, the nodes exhibit mixed block membership, potentially belonging to a different block for each tie considered. This model generalizes the usual stochastic block model (SBM, e.g., Snijders & Nowicki, 1997), in which each node belongs to one and only one block. Covariates may be added to the model by building GLM structures for the $\boldsymbol{\theta}_i$s or for the entries in B, for example.

2.2 Hierarchical Network Models

Turning now to our new framework for multiple-network models, we denote a collection of K networks as $\mathbb{Y} = (Y_1, \ldots, Y_K)$ where Y_k is an $n_k \times n_k$ tie matrix like that of Equation 1, for the n_k individuals in network k. Then, adding a network subscript k to the tie variables in Equation 1, Y_{ijk} is the value of the tie from individual i to individual j in network k. For example, in Figure 1, we have $K = 2$, $n_1 = 28$, and $n_2 = 27$. Once again, the tie values may be binary ($Y_{ijk} \in \{0, 1\}$), positive valued and unbounded ($Y_{ijk} \in [0, \infty)$), or may have some other specification.

To build our hierarchical Bayes generalization of the single network models in Section 2.1, we first write a likelihood, or general Level 1 model, for the ensemble Y of K networks,

$$P(\mathbb{Y}|\mathbb{X}, \Theta) = \prod_{k=1}^{K} P(Y_k|X_k = (X_{1k}, \ldots, X_{Pk}), \Theta_k = (\theta_{1k}, \ldots, \theta_{Qk})), \tag{5}$$

where $P(Y_k|X_k, \Theta_k)$ is one of the probability models in Section 2.1 (or indeed any other statistical model) for a single network Y_k with covariates X_k, and parameters Θ_k.

We then specify a very general Level 2 model for Θ_k as a hierarchical distribution F with hyperparameters ψ and other possible covariates W_k,

$$(\Theta_1, \ldots, \Theta_K) \sim F(\Theta_1, \ldots, \Theta_K | W_1, \ldots, W_K, \psi).$$

The underlying distribution F may be conceptualized as a super-population distribution from which particular networks are sampled, or as a Bayesian prior distribution and may be elaborated in various ways. For example, if F is specified so that

$$\Theta_k \overset{iid}{\sim} F^*(\Theta|\psi), k = 1, \ldots, K,$$

then the networks Y_k are exchangeable. Speaking generatively, Θ_ks are drawn iid from F^*, and then ties are generated randomly from the models specificed by the Θ_ks. Illustrations of the resulting networks are shown under the sets of ties for each of the K networks in Figure 2.

A key innovation to our approach is that we can model a dependence across networks by choice of W_k. Note that W_k may contain the same or different covariates across networks, providing a very general way to consider dependence across networks. Moreover, the presence of covariates W_k, and/or additional hierarchical distribution structure on the hyperparameters ψ, may lead to other structures, so that the networks are only partially exchangeable with one another (in contrast to the fully exchangeable structure of Figure 2).

This framework allows for a variety of dependence assumptions as well as a freer choice of social network model. In addition, our models accommodate covariates that may be network-, individual-, or tie-specific, or some combination. For example, X_k might contain tie-level covariates indicating that individuals i and j in network k are in the same (observable) group, individual-level covariates measuring teacher demographic information, and network-level covariates such as school proportion of free and reduced-price lunch status or the type of program or curriculum implemented.

Including a network-level group indicator allows us to model quantitative differences across groups of networks in experimental or observational data, as

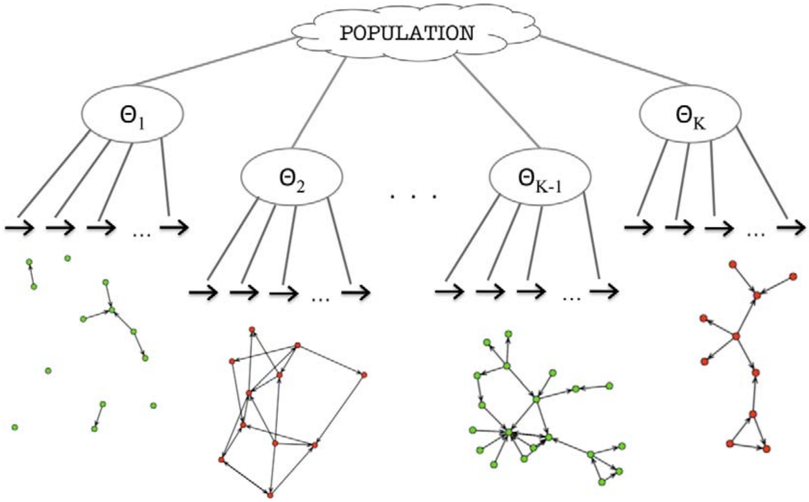

FIGURE 2. *A hierarchical structure illustrating a basic model for multiple networks in our framework. The parameters* $\Theta_k, k = 1, \ldots, K$ *are drawn independently from a common population, and then ties are generated from the model parameterized by each* Θ_k*, determining the structure of the* k*th network in the ensemble,* $k = 1, \ldots, K$.

well as treatment effects in an experimental context. Let T_k be the treatment or group indicator for network k and let α the treatment or group effect parameter. If there are only two treatments or groups, then T_k may be binary, and α a single real parameter. If there are multiple treatments or groups, then T_k may be a vector of indicators, and α the corresponding vector of effects for each treatment or group. Then a model for estimating the treatment effect/effects α may be written as

$$
\begin{aligned}
P(\mathbb{Y}|\mathbb{X}, \Theta, T, \alpha) &= \prod_{k=1}^{K} P(Y_k|X_k = (X_{1k}, .., X_{Pk}), T_k, \Theta_k = (\theta_{1k}, .., \theta_{Qk}), \alpha) \\
(\Theta_1, \ldots, \Theta_K) &\sim F(\Theta_1, \ldots, \Theta_K|\psi) \\
\alpha &\sim g(\alpha|\gamma) \\
&\quad \text{(with additional hierarchical structure, as appropriate).}
\end{aligned} \tag{6}
$$

This provides a general, partially exchangeable modeling framework for modeling multiple networks and estimating treatment effects.

3 HLSMs/Specializations of the HNM Framework

The hierarchical structure specified in Section 2.2 is extremely general, and many different types of network models, including those reviewed in Section 2.1, can be used to specify a particular HNM. Here, we show how the HLSM can be specified, before illustrating some examples of data analysis using the HLSM in Section 4.

Using the LSM likelihood in Equation 3 as the individual network model in the general structure in Equation 6, we may specify an HLSM as

$$P(\mathbb{Y}|\mathbb{X}, \beta, \mathbb{Z}, T, \alpha) = \prod_k \prod_{i \neq j} P(Y_{ijk}|Z_{ik}, Z_{jk}, X_{ijk}, \beta_k, T_k, \alpha), \tag{7}$$

where $P(Y_{ijk}|Z_{ik}, Z_{jk}, X_{ijk}, \beta_k, T_k, \alpha)$ is specified as

$$\operatorname{logit} P[Y_{ijk} = 1] = \beta_k^T X_{ijk} - d(Z_{ik}, Z_{jk}), \tag{8}$$

or in a particular experimental context,

$$\operatorname{logit} P[Y_{ijk} = 1] = \beta_k^T X_{ijk} - d(Z_{ik}, Z_{jk}) + \alpha T_k, \tag{9}$$

and we may supply additional hierarchical structure such as

$$\left.\begin{array}{rcr} Z_{ik} & \overset{iid}{\sim} & MVN(\mu_k, \Sigma_k) \\ \beta_{0k} & \sim & N(\upsilon_{0k}, \sigma_{0k}) \\ \vdots & & \\ \beta_{pk} & \sim & N(\upsilon_{pk}, \sigma_{pk}) \\ \alpha & \sim & N(\eta, \tau), \end{array}\right\} \tag{10}$$

where Z_{ik} is the latent position of individual i in the latent space for network k, X_{ijk} is a suitable set of covariates for individuals i and j in network k, with coefficients β_k, T_k is the treatment indicator ($T_k = 1$ if treated, $T_K = 0$ if not) for network k, and α is the treatment effect. Note that the prior distributions given are merely examples and any distribution may be used. Furthermore, additional hierarchical structure (linear or otherwise), as well as dependence on covariates at higher levels, may be imposed on the hyperparameters $\mu_k, \Sigma_k, \upsilon_{jk}, \sigma_{jk}, \eta$ and τ, as needed.

Note that T_k appears as a single additive treatment indicator here, so that α is the effect on overall tie probabilities (which will affect tie density) for treatment. However, T_k could be replaced by a vector of indicators for different treatment conditions, and/or could be placed elsewhere in the model, such as an interaction with a covariate X_{ijk}, a multiplicative effect on $d(Z_{ik}, Z_{jk}) = |Z_{ik} - Z_{jik}|$, and so forth.

We developed a Markov chain Monte Carlo (MCMC; Gelman, Carlin, Stern, & Rubin, 2004) algorithm, which we adapted from the single network algorithm described by Hoff, Raftery, and Handcock (2002), to fit the models in Equations 8 and 9 as well as similar models. Our algorithm uses Metropolis-Hastings updates to draw α, Z_k, β_k for each network k, and uses Gibbs updates for all other parameters, at each step of our MCMC algorithm.

Our MCMC algorithm also includes measures to constrain the configuration of latent space positions for each network. As detailed by Hoff et al. (2002), there is an identifiability issue with the latent space positions of the individuals. Since the model is defined based on pairwise distances, there are infinitely many possible configurations of teacher positions that produce the same set of pairwise distances. Our method differs from the postprocessing method described by Hoff et al. (2002) in that we constrain positions during the sampling process. We can uniquely identify a configuration of positions by imposing constraints to address translation, reflection, and rotation. For a two-dimensional latent space, for example, we fix the location of one individual to address translation, fix a coordinate of a second individual to address rotation, and fix the orientation of a third individual with respect to the first two individuals to address reflection. For higher dimensions, we can repeat the same constraints recursively. For n dimensions, we would constrain the first $n + 1$ positions; the first point is fixed, $n - 1$ dimensions of the second point are fixed, until we have only to fix 1 dimension of the $n + 1$ point. The R (R Development Core Team, 2011) source files for this algorithm are available on request from the authors.

The other standard network models described in Section 2.1 can be similarly incorporated into the HNM framework, replacing the likelihood portion of Equation 6 with the likelihoods in Equations 2 and 4, and adjusting the hierarchical structure accordingly, producing hierarchical ERGMs (HERGMs) and hierarchical MMSBMs (HMMSBMs), respectively. Of course, any other probabilistic model for a single network Y could be inserted as the likelihood in Equation 6 as well.

Finally, we note that HNMs are a natural extension of HLMs, that provide for different dependence structures in the lowest level clusters than HLMs do. Generalized linear HLMs could be fitted to tie data as in Figure 2, but the usual HLM framework requires independence among the units in the lowest level clusters, and this assumption is hardly ever appropriate for ties in social networks. Even though cross-classified HLMs (Raudenbush & Bryk, 2002) allow for some sender and receiver dependence, these models still assume some level of independence among ties that is inappropriate for many social network ties. For example, two friends of the same individual may be much more likely to be friends themselves, due perhaps to an unmeasured common interest that brought them together, the fact that they spend more time together because of their common friend, and so on. These dependence relations can be accounted for within the lowest level clusters in an HNM (see Figure 2). In particular, in an HLSM, if individuals i and j are both friends with individual ℓ, then the latent positions Z_i and Z_j will likely be close to Z_ℓ; hence, Z_i and Z_j will be relatively close to each other, and the likelihood of a friendship tie between i and j will be relatively greater than between other pairs of individuals. By contrast, an HLM for ties structured like Figure 2 assumes independence among ties in each network, and cannot capture this extra "friends of a friend" dependence between ties.

4 Empirical Examples

To demonstrate that our approach can detect meaningful effects in education network data, we fit HLSMs to two network data sets. The first data set is

TABLE 1
Pitts and Spillane Descriptive Statistics by Network

Network Size	Density	Proportional Density	Mean In-/Out-Degree	Proportion Teaching the Same Grade
30	61	0.07	2.03	0.60
21	103	0.25	4.90	0.38
25	75	0.13	3.00	0
19	54	0.16	2.84	0.11
26	76	0.12	2.92	0.65
49	162	0.07	3.31	0.53
76	314	0.06	4.13	0.61
42	172	0.10	4.10	0.64
33	106	0.10	3.21	0.76
29	107	0.13	3.69	0.41
12	15	0.11	1.25	0.17
28	117	0.15	4.18	0.50
14	46	0.25	3.29	0.14
14	51	0.28	3.64	0.43
15	81	0.39	5.4	0.40

Note. Network Size is the number of nodes; Density is the total number of ties; Proportional Density is the proportion of possible ties. Mean In-/Out-Degree is the mean number of ties sent or received in each network, and Proportion Teaching the Same Grade is the proportion of teachers who is teaching the same grade as at least one other teacher. Note that teachers who are missing grade taught are considered not to teach the same grade as any other teacher.

composed of teachers whose ties indicate the seeking of professional advice, collected by Pitts and Spillane (2009); our application shows that the model can detect the effect of tie-level covariates. The second data set is simulated data, which we use to explore the utility and operating characteristics for detecting network-level treatment effects.

4.1 Covariate Effects in an Observational Study

Pitts and Spillance (2009) surveyed teachers and principals in 15 elementary and middle schools from a large, urban school district to validate a school staff survey instrument. These schools include both private and public schools ranging from pre-kindergarten to eighth grade and vary in the number of staff members. School staff were asked to list to whom they seek advice and include the frequency and value they place on the advisor. Additional demographic data and belief measures were also collected.

Networks consist of teachers who responded to the network portion of the survey and other teachers identified as tie recipients. Schools varied in network size, the smallest network of 12 teachers and the largest of 76 teachers. Table 1 lists network size and other descriptive network statistics for each school. Smaller schools tend to have a higher number of ties relative to network size whereas schools tend to have similar numbers of ties per teacher, regardless of size. In addition to network statistics, we include in Table 1 one edge-level variable, the proportion of teachers who taught the same grade as at least 1 other teacher. Although schools include subsets of grades pre-k through 8, this proportion varies widely by school. This is due to missing data since teachers who did not report their grade assignment cannot be matched. There are four schools for which missing data is so extensive that less than 20% of the teachers are teaching the same grade as another teacher.

For these data, we fit a simple HLSM with a single tie-specific covariate; $X_{1ijk} = 1$ if teacher i and j in school k teach the same grade and 0 otherwise. For simplicity, teachers who did not report their grade taught are assumed not to teach the same grade as any other teacher.

We specify the model as

$$\begin{aligned}
\operatorname{logit} P[Y_{ijk} = 1] = \;& \beta_{0k} + \beta_{1k} X_{1ijk} - |Z_{ik} - Z_{jk}|,\\
Z_{ik} &\sim MVN\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\ 0 & 10\end{pmatrix}\right), i = 1, \ldots, n_k,\\
\beta_{0k} &\sim N(\mu_0, \sigma_0^2), k = 1, \ldots, K,\\
\beta_{1k} &\sim N(\mu_1, \sigma_1^2), k = 1, \ldots, K,\\
\mu_0 &\sim N(0, 1),\\
\mu_1 &\sim N(0, 1),\\
\sigma_0^2 &\sim \mathrm{Inv-Gamma}(100, 150),\\
\sigma_1^2 &\sim \mathrm{Inv-Gamma}(100, 150),
\end{aligned} \tag{11}$$

where $K = 15$, the total number of schools. We model the latent space positions for teachers in each school as two-dimensional. Furthermore, we impose an additional hierarchical structure by estimating the hyperparameters μ_0, μ_1, σ_0^2, σ_1^2.

We fit our model using the MCMC algorithm sketched in Section 3. The Metropolis-Hastings draws were tuned to ensure acceptance rates between 0.38 and 0.46 for all one-dimensional parameters and between 0.2 and 0.28 for parameters with higher dimensions (Gelman et al., 2004). Visual inspections of trace plots and running means plots suggest good mixing and short burn-in lengths of less than 10,000 steps (Figure 3). Autocorrelation plots suggest thinning every 50 steps (Figure 4).

We removed the first 10,000 steps as burn-in and thinned the resulting chain; every 50th step was retained, for a total posterior sample of 1,801 steps. Using this sample, we plotted posterior medians along with 50% and 95% equal-tailed credible intervals for β_0 and β_1 for each school in Figure 5.

There is wide variability in intercept and slope posterior medians for each school. Although schools vary in their overall probability of making a tie (β_0), there seems to be a positive effect of teaching the same grade (β_1). We model β_0 and β_1 from normal distributions, see Section 4.1. Posterior means are (1.50, 1.47) and (1.85, 4.34) for (μ_0, σ_0^2) and (μ_1, σ_1^2), respectively. Moreover, the 95% credible interval for μ_1 is (0.77, 2.93), so the global effect of teaching the same grade is positive.

There are several possible explanations for the variability in β_0 across schools. Some variability is expected, as schools vary in the social structure of their teachers, but there are other possible explanations as well. There are several schools, Schools 6 through 10 and 12, which have much larger networks than the other schools. The intercept for these schools would naturally be a bit lower, since the overall probability of a tie between two teachers would decrease, as the number of individuals within a school increases. In fact, the schools with large positive intercepts are the schools smallest in size, Schools 11, 13, 14, and 15 with 15 or fewer teachers. These schools also happen to be parochial schools.

Regarding β_1, the effect of teaching the same grade, we expect the size of the school to be correlated with the coefficient estimates since schools with many teachers teaching the same grade are less likely to have mutual ties than schools with only a few teachers teaching the same grade. Rather, what we find is the size of the school, coupled with the amount of grade assignment data available, drives the variability in the β_1 estimates. Large schools tend to have more data available and the four largest schools, Schools 6 through 9, generally have small variability in their estimates for β_1. Schools that are either small or have little information regarding teacher grade assignment have large variability; Schools 3, 4, 11, and 13 have relatively little information, and Schools 11, 13, 14, and 15 have the fewest numbers of teachers. For comparison, we also fit separate single-network LSMs for each of the 15 schools. The posterior spreads for β_1 are unacceptably large for several of the schools. Schools 3, 11, 13, and 15 have β_1 posterior variances of 104, 54, 41, and 32, respectively.

4.2 Treatment Effects in a Controlled Experiment

We do not yet have social network data involving a controlled, network-level intervention. However, studies like Spillane, Correnti, and Junker (2009) offer the promise of such data in the near future. In the meantime, we have simulated several social network data sets, to explore the utility and operating characteristics of our models for detecting treatment effects.

Each simulated data set consists of 20 schools, with 10 teachers in each school. We generated undirected, binary network ties from the model

$$\operatorname{logit} P[Y_{ijk} = 1] = 2 + 4X_{1ijk} - |Z_{ik} - Z_{jk}| + \alpha T_k, \tag{12}$$

where half of the schools are assigned to be treatment schools ($T_k = 1$) and the other half are control schools. We included only one other covariate, the tie-level binary indicator that teacher i and teacher j teach the same grade. In the

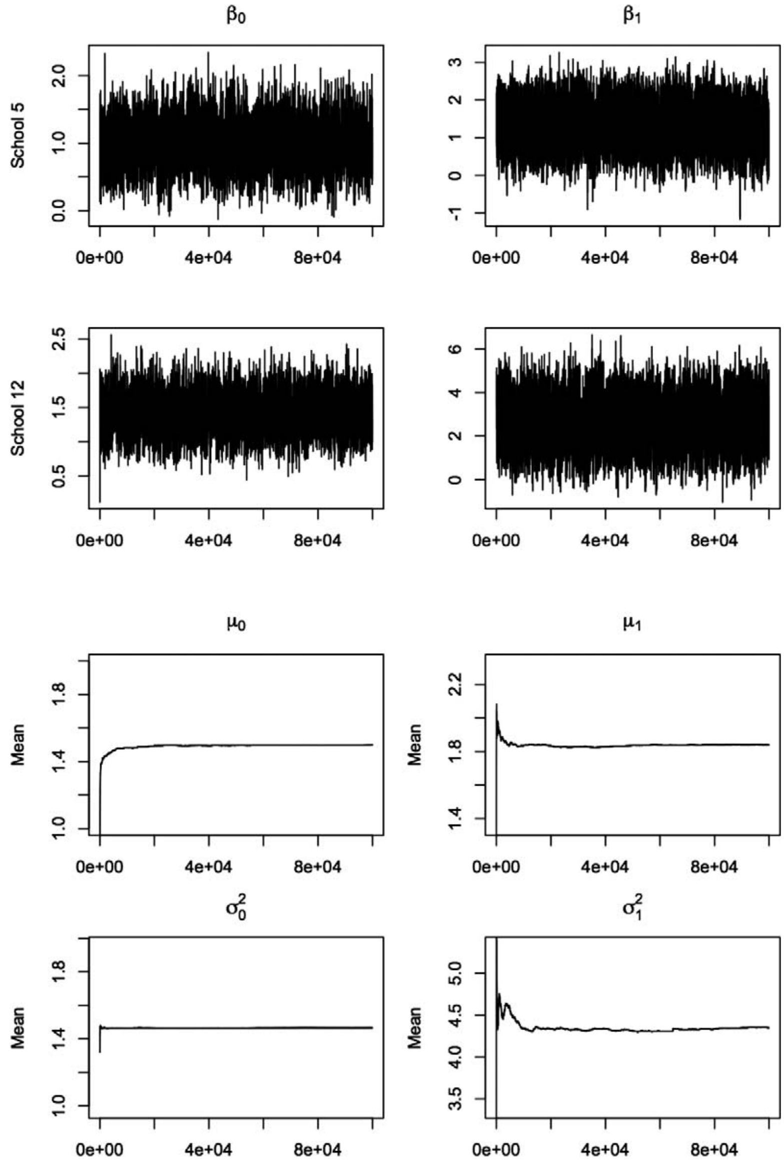

FIGURE 3. *Trace plots and running means diagnostics plots for the hierarchical latent space model applied to the data of Pitts and Spillane (2009), Schools (5, 12) and hyper-parameters* μ *and* σ^2.

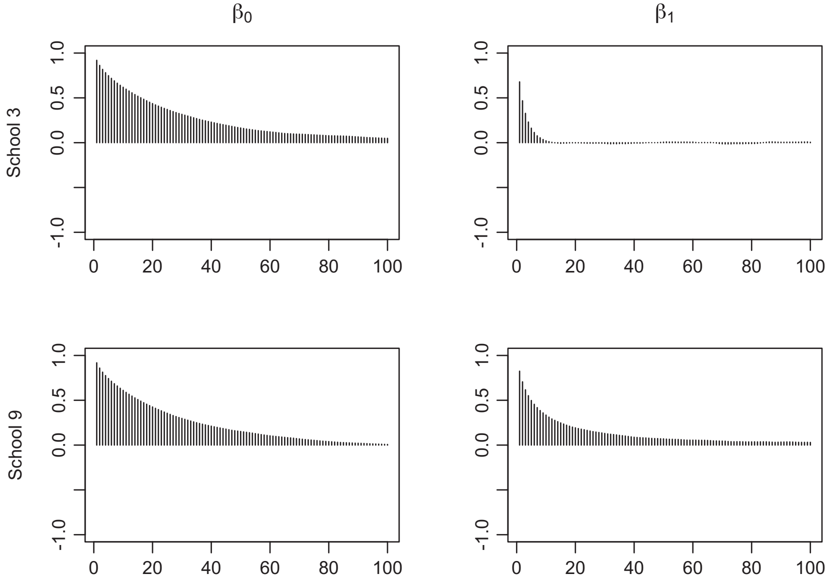

FIGURE 4. *Autocorrelation diagnostics plots for the hierarchical latent space model applied to the data of Pitts and Spillane (2009), Schools (3, 9) suggest retaining every 50 steps.*

simulation, teachers were randomly assigned to one of four grade levels, and teacher latent space positions were chosen deliberately to cover a range of pairwise distances that yield a range of contributions to the probability of a tie. We simulated four sets of data, one for each treatment effect value of $\alpha = 3,\ 2,\ 1,\ 0.5$; several of these networks are shown in Figure 6.

Given the data generating model (Equation 12), the HLSM we fit has a similar structure,

$$\begin{aligned} \operatorname{logit} P[Y_{ijk} = 1] &= \beta_0 + \beta_{1k} X_{1ijk} - |Z_{ik} - Z_{jk}| + \alpha T_k, \\ Z_{ik} &\sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}\right), i = 1, \ldots, n_k, \\ \beta_0 &\sim N(0, 10), \\ \beta_{1k} &\sim N(0, 10), k = 1, \ldots, K, \\ \alpha &\sim N(0, 10), \end{aligned} \tag{13}$$

where $K = 20$, the total number of schools. We model the latent space positions for teachers in each school as two-dimensional. We treat the intercept β_0 as fixed and constant across all schools but allow the slopes β_{1k} to vary by school. Specifying very diffuse priors for α, β_0, and β_{1k} is essentially equivalent to treating them as fixed effects in the estimation.

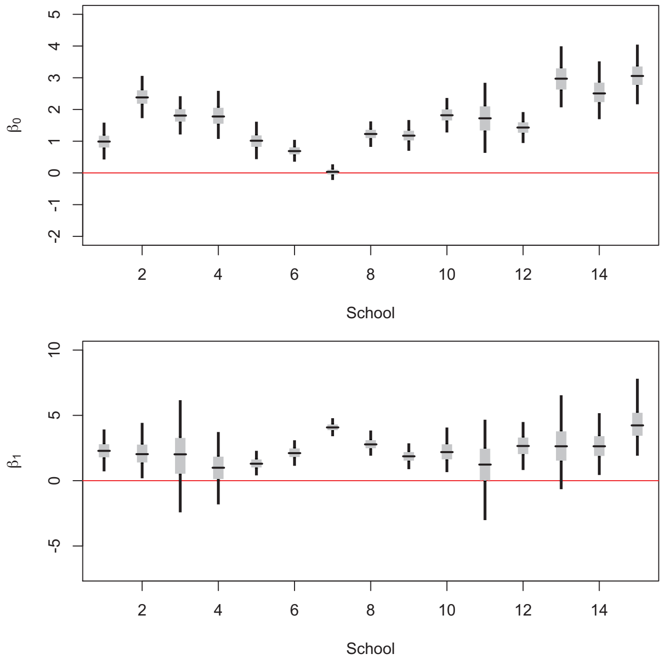

FIGURE 5. *Posterior modes (horizontal black line segments) and 95% intervals (vertical black line segments) for the intercept (β_0) and slope (β_1) parameters in the hierarchical latent space model fitted to the data of Pitts and Spillane (2009). Most 95% intervals for β_1 are completely or mostly above zero suggests that teachers have a greater tendency to seek advice from other teachers in the same grade than across grades, controlling for other factors using the latent space for each network.*

We fit this model using the MCMC algorithm sketched in Section 3, for each simulated data set. Metropolis-Hastings draws were tuned so that the acceptance rates were between 0.38 and 0.48 for all one-dimensional parameters and between 0.2 and 0.28 for the two-dimensional latent space positions (Gelman et al., 2004). Although traceplots suggest convergence early on, running means plots suggest longer burn-in is needed (Figure 7). Furthermore, autocorrelation plots suggest high levels of correlation (Figure 8). Therefore, we used a burn-in period of 30,000 iterations and retained 1 of every 150 steps, for a total posterior sample of 401 steps.

Using these samples, we plot posterior distributions for the treatment effect α for each simulated data set. Figure 9 also displays the true value of α as a vertical line, and equal-tailed 95% credible intervals are inferred by the shaded region.

For each posterior sample, the associated 95% credible interval covers the true value of α indicating accurate parameter estimation. Two of the posterior modes

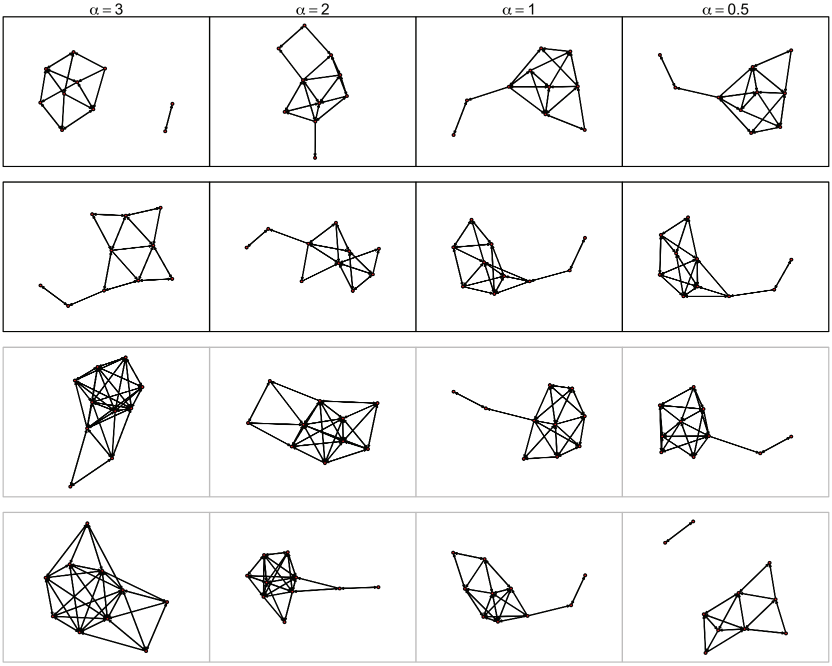

FIGURE 6. *Networks shown for two schools in each condition simulated for each value of $\alpha = 3, 2, 1, 0.5$. The top two rows show the control condition networks and the bottom two rows show the treatment condition networks. For higher values of α, the treatment condition networks are noticeably more dense than the control condition networks.*

are less than the true value of α and the other two are slightly greater than the true value of α. We attribute this to measurement error since our network tie data is binary. And in fact, other versions of this simulation study yield posterior modes for α were either less than or very close to the true value for all of the simulated data. We also calculated the correlation between the true pairwise latent space distances and the fitted pairwise latent space distances and find that 80% of the fitted models have correlations > 0.72 and 70% have correlations > 0.81, suggesting good recovery of pairwise distances.

In experiments, the ability to detect existing treatment effects is of great importance and we can use our 95% credible intervals to determine significant treatment effects (Figure 9). We recover significant treatment effects from the simulated data when the true α is 3, 2, or 1, despite a small sample of networks and a small number of teachers within each school. Significant treatment effects are not recovered when $\alpha = 0.5$. This is not particularly surprising, as a 0.5 increase in log odds translates into a minimal increase in tie probability for most probability values.

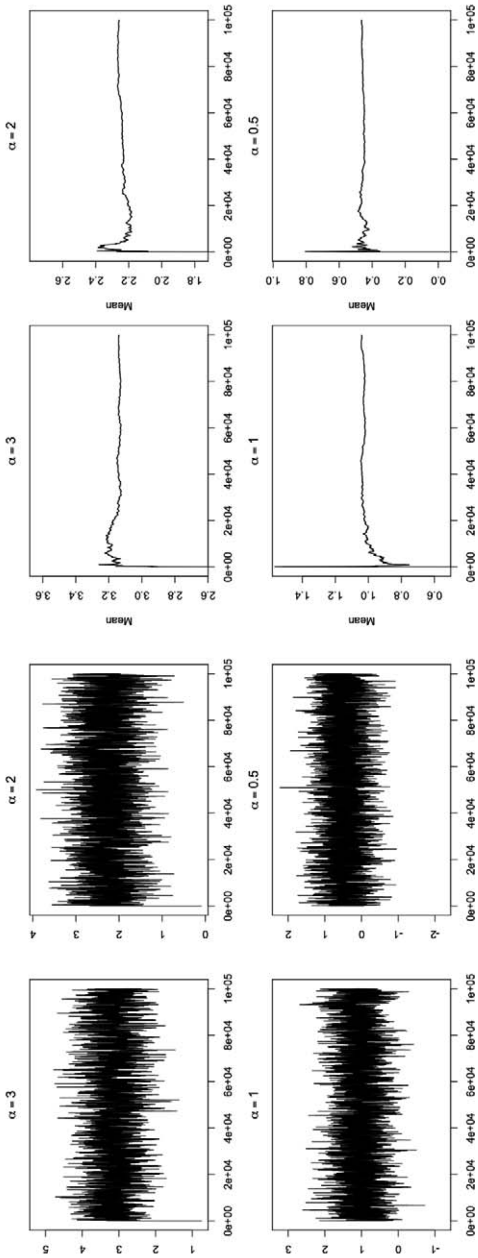

FIGURE 7. *Trace plots and running means diagnostic plots of the treatment effect α for each simulation. While trace plots suggest early convergence, running means plots suggest a longer burn-in is needed.*

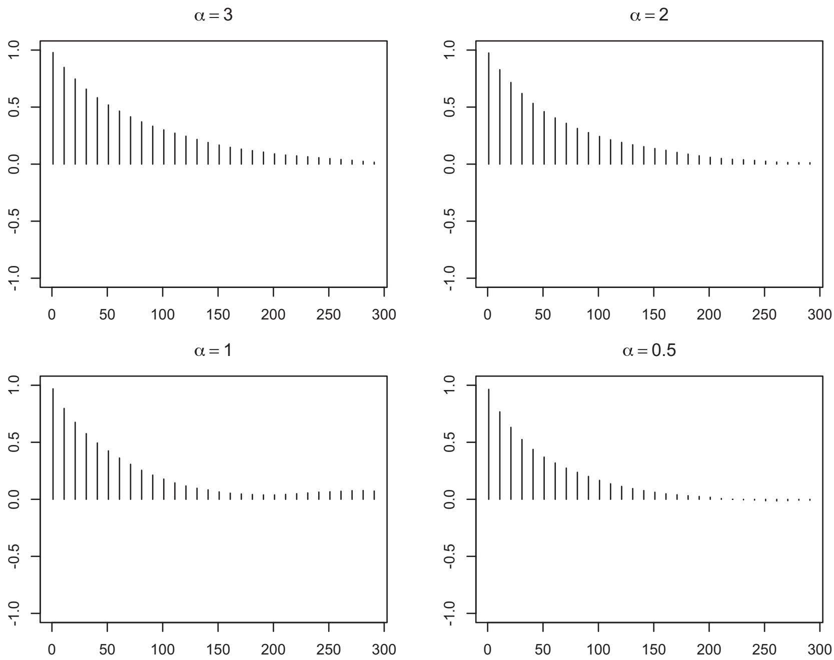

FIGURE 8. *Autocorrelation diagnostic plots of the treatment effect α for each simulation, suggesting that a large amount of thinning is required for uncorrelated draws.*

5 Discussion

Many current theories of action (e.g., Glennan & Resnick, 2004; Hord et al., 2010; McLaughlin & Talbert, 2006) and proposed or implemented interventions (e.g., Matsumura et al., 2010; Spillane et al., 2009) in education research focus on the network of relationships among teachers in school buildings, students in classes, and so on. In order to efficiently and accurately judge the effect of such interventions, we need statistical social network models that can accomodate multiple partially exchangeable networks, as well as treatment effects and other covariate effects on network structure. Current social network analysis is largely inadequate for this task, focusing either on descriptive statistics for the networks or full statistical models restricted to one network at a time.

To address the special needs of social network analysis in education research, we have introduced the HNM framework. Models developed in this framework can accommodate relational data from several networks and can be parameterized to model treatment effects in a variety of ways. HNMs extend single-network models using multilevel structure analogous to HLMs, replacing the usual conditional independence structure at Level 1 with dependence structures appropriate for network modeling, and allowing essentially arbitrary modeling at higher levels.

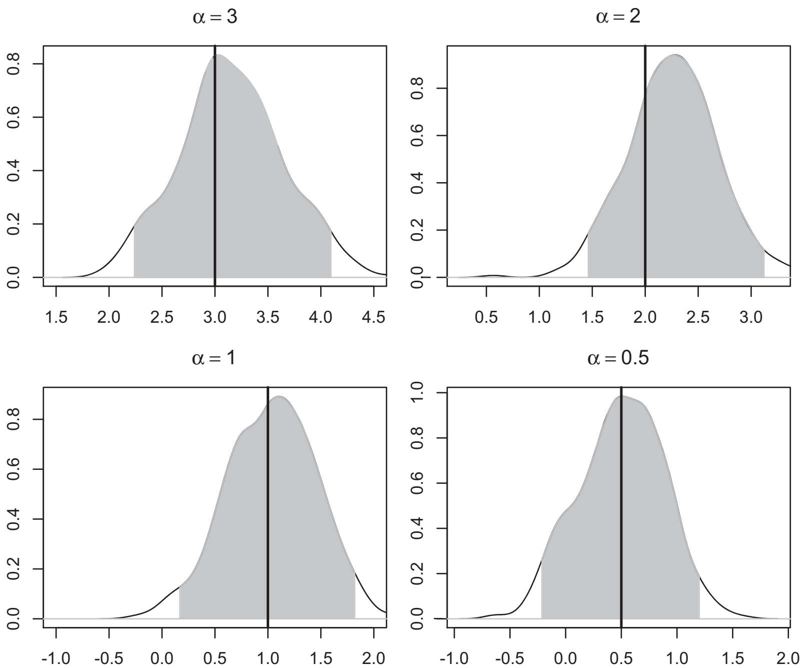

FIGURE 9. *Posterior density estimates for different values of the true treatment effect (α) for each simulation. Gray regions indicate 95% credible intervals; the black vertical lines represent the true treatment effect in each simulation.*

We illustrated the HNM framework by extending single-network latent space models (Hoff et al., 2002) to HLSMs, and introduced an MCMC algorithm to fit HLSMs. We introduced a simple set of constraints to address potential configuration identifiability problems in two dimensions, and indicated an extension to higher dimensions. We also showed that the HLSM family of models can successfully detect covariate and treatment effects in realistically sized observational and experimental studies in education research.

This article serves as a "proof of concept" for the HNM framework. A key issue for future work in this area is higher dimensional latent spaces. Our identifiability constraints should be compared with the postprocessing methods introduced by Hoff et al. (2002). The choice of dimension is another area for future work, since it is unknown whether higher dimensions are necessary.

The HNM framework is larger than latent space models alone; we are also extending the MMSBM (Sweet, Thomas, & Junker, 2012) to the HNM setting. Future work includes extending the work done by Zijlstra et al. (2006) and Templin et al. (2003) to formally adapt ERGMs to the HNM setting, and developing model criticism and model selection methods for these models.

The HNM framework is well suited to estimating treatment and covariate effects in multiple-network experiments. Often, however, changes in the social network are intermediate variables for an intervention whose intended "final" outcome is student achievement or a similar variable (e.g., McLaughlin & Talbert, 2006). Since the HNM is itself a multilevel model, it can be emebedded as the intermediate level of a multilevel model for the outcome of interest—in this way, social network structure need not be ignored in studies of other educational outcomes of interest.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Program for Interdisciplinary Education Research, supported by the Institute for Education Sciences, Department of Education, Grant R305B040063. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.

Note

1. For example, if we measure not only the number of contacts between teacher i and teacher j but also the average length of time of a contact, then each Y_{ij} would be an ordered pair of these values.

References

Airoldi, E., Blei, D., Fienberg, S., & Xing, E. (2008). Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research*, *9*, 1981–2014.

Bonsignore, E., Hansen, D., Galyardt, A., Aleahmad, T., & Hargadon, S. (2011). The power of social networking for professional development. In T. Gray & H. Silver-Pacuilla (Eds.), *Breakthrough teaching and learning: How educational and assistive technologies are driving innovation* (pp. 25–52). New York, NY: Springer.

Coburn, C., & Russell, J. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, *30*, 203–235.

Daly, A., Moolenaar, N., Bolivar, J., & Burke, P. (2010). Relationships in reform: The role of teachers' social networks. *Journal of Educational Administration*, *48*, 359–391.

Fienberg, S., Meyer, M., & Wasserman, S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, *80*, 51–67.

Frank, K. A., Penuel, W. R., Sun, M., Chong, M. K., & Singleton, C. (2013). The organization as a filter of institutional diffusion. *Teachers College Record, 115*(1).

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education*, *77*, 148–171.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. Boca Raton, FL: CRC press.

Glennan, T. K., & Resnick, L. B. (2004). Expanding the reach of education reforms perspectives from leaders in the scale-up of educational interventions. In T. K. Glennan, S. J. Bodill, J. R. Galegher, & K. A. Kerr (Eds.), *School districts as learning organizations: A strategy for scaling education reform* (*Chap. 14*, pp. 517–563). Santa Monica, CA: RAND Corporation.

Goldenberg, A., Zheng, A., Fienberg, S., & Airoldi, E. (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning*, *2*, 129–133.

Harris, K., Halpern, C., Whitsel, E., Hussey, J., Tabor, J., Entzel, P., & Udry, J. (2009). *The national longitudinal study of adolescent health: Research design*. Retrieved from http://www.cpc.unc.edu/projects/addhealth/design

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.

Hord, S., Roussin, J., & Sommers, W. (2010). *Guiding professional learning communities: Inspiration, challenge, surprise and meaning*. Thousand Oaks, CA: Corwin/Sage.

Johnson, M., & Junker, B. (2003). Using data augmentation and Markov chain Monte Carlo for the estimation of unfolding response models. *Journal of Educational and Behavioral Statistics*, *28*, 195.

Kolaczyk, E. (2009). *Statistical analysis of network data: Methods and models*. New York, NY: Springer-Verlag.

Lazega, E., & van Duijn, M. (1997). Position in formal structure, personal characteristics and choices of advisors in a law firm: A logistic regression model for dyadic network data. *Social Networks*, *19*, 375–397.

Lin, N. (1999). Building a network theory of social capital. *Connections*, *22*, 28–51.

Matsumura, L., Garnier, H., & Resnick, L. (2010). Implementing literacy coaching: The role of school social resources. *Educational Evaluation and Policy Analysis*, *32*, 249–272.

McLaughlin, M., & Talbert, J. (2006). *Building school-based teacher learning communities: Professional strategies to improve student achievement (Vol. 45)*. New York, NY: Teachers College Press.

Moreno, J. (1934). *Who shall survive?: A new approach to the problem of human interrelations*. Washington, DC: Nervous and Mental Disease.

Moolenaar, N., Daly, A., & Sleegers, P. (2010). Occupying the principal position: Examining relationships between transformational leadership, social network position, and schools' innovative climate. *Educational Administration Quarterly*, *46*, 623.

Penuel, W., Riel, M., Joshi, A., Pearlman, L., Kim, C., & Frank, K. (2010). The alignment of the informal and formal organizational supports for reform: Implications for improving teaching in schools. *Educational Administration Quarterly*, *46*, 57–95.

Penuel, W., Riel, M., Krause, A., & Frank, K. (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *The Teachers College Record*, *111*, 124–163.

Penuel, W. R., Frank, K. A., & Krause, A. (2006). The distribution of resources and expertise and the implementation of schoolwide reform initiatives. In *Proceedings of the 7th international conference on Learning sciences*, ICLS '06 (pp. 522–528). Indiana University, Bloomington, IN: International Society of the Learning Sciences.

Pitts, V., & Spillane, J. (2009). Using social network methods to study school leadership. *International Journal of Research & Method in Education*, *32*, 185–207.

Polak, M., Heiser, W., De Rooij, M., & Busing, F. (July 2003). A comparison of correspondence analysis, multidimensional unfolding and the generalized graded unfolding model for single-peaked data. Paper presented at the 13th International Meeting of the Psychometric Society Chia Laguna (Cagliari), Italy.

R Development Core Team. (2011). *R: A language for data analysis and graphics*. Vienna, Austria: R Foundation for Statistical Computing.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods (Vol. 1).* Thousand Oaks, CA: Sage.

Snijders, T., & Kenny, D. (1999). The social relations model for family data: A multilevel approach. *Personal Relationships*, *6*, 471–486.

Snijders, T., & Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, *14*, 75–100.

Spillane, J., Correnti, R., & Junker, B. (2009). *Learning leadership: Kernel routines for instructional improvement. (IES Grant Proposal.)* Chicago, IL: Author.

Sweet, T., Thomas, A. C., & Junker, B. (2012). *Handbook on mixed membership models*, chapter Hierarchical Mixed Membership Stochastic Blockmodels for Multiple Networks and Experimental Interventions. New York, NY: Chapman & Hall/CRC. invited chapter.

Templin, J., Ho, M.-H., Anderson, C., & Wasserman, S. (2003). Mixed effects p* model for multiple social networks. In Proceedings of the American Statistical Association: Bayesian Statistical Sciences Section, (pp. 4198-4024). Alexandria, VA: American Statistical Association.

Thomas, S. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *The Journal of Higher Education*, *71*, 591–615.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, *61*, 401–425. doi:10.1007/BF02294547

Weinbaum, E., Cole, R., Weiss, M., & Supovitz, J. (2008). Going with the flow: Communication and reform in high schools. In J. Supovitz & E. Weinbaum (Eds.), *The implementation gap: Understanding reform in high schools* (pp. 68–102). New York, NY: Teachers College Press.

Zijlstra, B., van Duijn, M., & Snijders, T. (2006). The multilevel p 2 model. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *2*, 42–47.

Authors

TRACY M. SWEET is a post-doctoral fellow in the Statistics department at Carnegie Mellon University, Pittsburgh, PA 15213; tracymsweet@gmail.com. Her research interests include education statistics, and in particular, research methodology for large-scale experiments and social network analysis of professional communities.

ANDREW C. THOMAS is a visiting assistant professor in the Statistics department at Carnegie Mellon University, Pittsburgh, PA 15213; acthomas@stat.cmu.edu. In addition to his research on social network analysis, his research interests include statistics in the social and political sciences, sports statistics, and biostatistics.

BRIAN W. JUNKER is a professor in the Statistics department at Carnegie Mellon University, Pittsburgh, PA 15213; brian@stat.cmu.edu. His research interests include nonparametric and Bayesian item response theory and unfolding models, hierarchical models for multiple ratings of extended-response test items, psychometric cognitive diagnosis models and predictive modeling of online tutoring systems, multiple-recapture census systems, mixed effects models and sampling weights in large-scale surveys, causal modeling, Markov chain Monte Carlo and other computing and estimation methods, errors in variables, factor analysis and structural equation models in econometrics and psychiatric statistics, statistical analysis of large randomized field trials in education, rating protocols for teacher quality, educational data mining, social network analysis, and grade of membership or latent dirichlet allocation models.

Manuscript received December 8, 2011
Revision received March 22, 2012
Accepted June 7, 2012