Temporal Latent Space Network Model with Covariates for Applications in Education Research

Samrachana Adhikari, Brian Junker *
Department of Statistics, Carnegie Mellon University

January 20, 2016

Abstract

In this paper, we extend existing latent space approach for network analysis to further account for temporal dependencies among networks. Motivation for this model is to understand and perform statistical analysis on how the structures of the network as represented by the latent positions evolve with time. Further in many real world applications, it is also of interest to understand the changing effects of the observed nodal covariates in tie formation. To address these questions, we introduce a Bayesian longitudinal latent space network model by combining ideas from state space model literature and latent variable network models. We leverage on the dual interpretation of the latent distance as latent variable and as residual to include edge covariates in the model. We then apply the proposed model to analyze temporal network of teachers' interactions in several secondary schools in a school district.

Keywords: Latent space model, Network attributes, Longitudinal network analysis, Advice seeking network of teachers

*The authors gratefully acknowledge Professor Jim Spillane from the Distributed Learning Study at NorthWestern University for providing data and valuable insights for this work.

1 Motivation and Introduction

Network data is widely used in education research to study how different units, such as teachers, students, administrators etc., in educational settings interact among themselves or with other units, and the implications of these interactions on the overall learning behavior of the students. Pitts and Spillane (2009), Spillane et al. (2012), Spillane and Hopkins (2013), Penuel et al. (2009) are some of the examples of such studies. Statistical analysis of educational networks helps us to evaluate teachers and students by identifying important properties and structures of the networks that affect their performances, and hence to make necessary policy recommendations for any improvement (Sweet et al., 2013; Sweet, Thomas, and Junker, Sweet et al.). Further, since interaction itself is a dynamic process that changes over time, conditional on external and internal factors, it is important to study the dynamics of the changing structures in networks observed over many time points. While majority of the earlier works in education literature have mainly focussed on networks observed at a particular time point, recently there has been increasing interest in studying networks over multiple time points (Spillane et al., 2012; Harris et al., 2013). However, development in methodologies is lagging compared to the interest generated in application and data collection. The changing nature and the added complexity of the data requires more sophisticated method of analysis to make sound inference on the data.

Consider, for example, instructional advice seeking network of teachers in a school district observed in 5 different years (Spillane et al., 2012). The socioplots of the observed advice seeking networks are shown in Figure 1. Number of changes were made in the schools since the first year of data collection. Assessment was changed from district level to state level, first for Reading and then for Math in the following year; new math coaches were introduced in the schools after the first year of data collection. We expect these changes to change teachers' incentives for communication and also their expectations from each other in terms of instructional advice. It is then of interest to understand how advice seeking behavior of the teachers as shown by the network is changing with time (Spillane et al., 2012). Further, researchers in education also want to understand how similarities or differences in observed covariates of the nodes or group of nodes affect present and future advice seeking behavior of teachers in the network. For example, Spillane et al. (2012) have shown that the teachers who teach in the same grade are more likely to seek advice from each other compared to those teaching in different grades. It is then natural to ask if the relationship between the covariates and the advice seeking behavior of the teachers is persistent over time in a longitudinal setting. This question has mostly remained unexplored in education research. It will also be useful to predict how the network will look like in the future, based on the past networks and covariates, to make important decisions in the schools.

In this paper, we introduce a method to analyze networks observed over several time points, and investigate how the relationship between network and observed covariates is changing over time in the presence of latent variables in longitudianl advice seeking network of teachers. We apply this method to stu We focus on extending existing method for network analysis, latent space network model to be more specific, that was introduced by Hoff et al. (2002) to model the evolution of the network structure over time. In the latent space network model, the underlying network structure is represented by the positions of nodes in a continuous (Euclidean) latent space. This class of model allows for the basic network properties like reciprocity and transitivity of the nodes, with possible extension to clusterability. We aim to extend the idea of reciprocity and transitivity in a static network to the temporal network settings. The nodes with ties at previous time points are more likely to have ties in the future, indicating that they will lie close to each other in the future

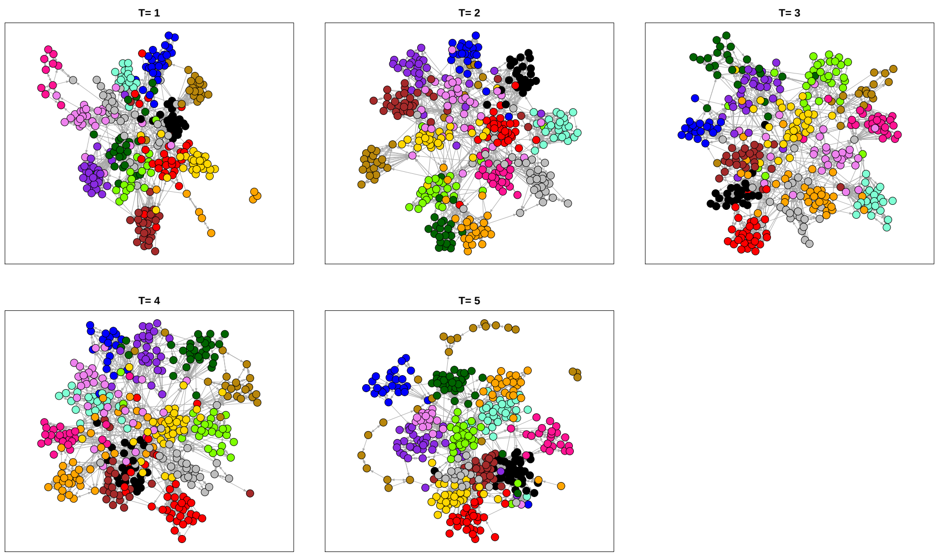


Figure 1: Socioplot of the advice seeking network of teachers in secondary schools in a school district (Spillane et al., 2012). Nodes are the teachers and a tie from i to j is defined by whether teacher i goes to j for advice. Different colors in a network represent different schools in the district.

latent space. Similary, if the nodes i and j, and the nodes i and k have ties at time t, the nodes i and k are more likely to have a tie at $t + 1$. Furthermore, a network can evolve in time by the expansion of the latent positions making the network sparser, by the shrinkage of the latent positions which makes the network denser, or by no change in the latent positions which implies no substantive change in the network structure, after accounting for the observed covariate effects in the network. Methods that treat related networks to be independent over time are not well-suited for networks observed over multiple time points (as we show later in Figure 6). By modeling temporal networks as different instances of static networks, we lose important information about network evolution. Instead, modeling them as a related process that accounts for the time dependencies helps us understand the evolution of the underlying network structure.

We propose a state space modeling approach for evolution of temporal networks using a class of latent variable models. By leveraging on the dual interpretation of latent variable and residual in the model, we furthur introduce a way to include and evaluate the effect of covariates on the likelihood of forming network ties over time. The proposed model is a crucial contribution to applications in education and other social-sciences, as it aims to answer some of the substantial questions, for instance pertaining to interactions of teachers observed over time, via a systematic statistical approach.

The organization of the paper is as follows. We review existing literature for analyzing longitudinal network in section 2. We present the proposed method in section 3 and the estimation of the model parameters using MCMC in section 4. We demonstrate the use of the model to analyze advice seeking network of teachers and the results in section 5, and end with summary and conclusion in section 6.

2 Literature Review

Existing statistical models for analyzing network observed at a time point can be broadly divided into two classes. (See Goldenberg et al. (2010), Fienberg (2012) and Dabbs et al. (prep) for a detailed review of these methods.) The first class of method is based on ***exponential random graph models (ERGMs)***, which are defined by different descriptive network statistics (Frank and Strauss, 1986). The probability distribution of a graph in ERGM can be written as an exponential family with the network statistics being the sufficient statistics for the distribution (Holland and Leinhardt, 1981; Robins et al., 2007; Goldenberg et al., 2010). The other class of method is based on ***latent variable or conditionally independent dyads (CID) models*** which are developed to model the underlying unobserved network features. Conditional on these latent features, ties in a network are assumed to be independent (Goldenberg et al., 2010; Dabbs et al., prep). Sender-receiver network model, latent space network model, stochastic block model and mixed membership stochastic block model are examples of the latent variable model (Hoff and Ward, 2004; Hoff et al., 2002; Airoldi et al., 2006). Further, Dabbs et al. (prep) have recently proposed an additive model that includes different combination of these latent variable models along with covariates.

Recently, there has been some works in statistics literature on extending existing network methods to account for temporal networks, for example by Robins and Pattison (2001), Hanneke and Xing (2007),Hanneke et al. (2010), Westveld and Hoff (2011), Xing et al. (2010), Sarkar and Moore (2005) and Sewell and Chen (2014). Robins and Pattison (2001), Hanneke and Xing (2007) and Hanneke et al. (2010) have studied the networks observed over discrete time points in the ERGM settings, also known as ***temporal ERGMs or TERGMs***. TERGMs make standard Markov assumption on the evolution of a network graph such that the network observed at time t is independent of networks observed in the previous time points un to $t-2$, given the network at $t-1$ (Hanneke and Xing, 2007), with an additional assumption that the distribution of the network observed at time t conditional on the network at $t-1$ has an ERGM representation (Frank and Strauss, 1986). As the assumption on tie dependence deviates from independence, the model gets more and more complicated. Snijders (1996) have also developed stochastic actor oriented models using a continuous time Markov processes, the class of models that is very similar to ERGMs. Many of these methods run into computational complications and convergence issues, for example convergence to degenerate solutions, when the model specification is complex, thus making them less desirable in real world applications (O'Malley, 2013; Handcock et al., 2003).

Westveld and Hoff (2011) extended the static model for directed networks with sender and receiver random effects and fixed covariate effects, introduced by Gill and Swartz (2001) and later implemented by Hoff and Ward (2004), to account for temporal dependencies in ***mixed effects temporal model***. They assume autoregressive dependence structure on the sender-receiver effects and the overall residual, and hence account for additional correlation in random error (random effects) introduced by temporal dependencies. Xing et al. (2010) extended ***mixed membership stochastic block model (MMSBM)*** (Airoldi et al., 2006) to account for the temporal nature of networks, and called it the ***dynamic mixed membership stochastic block model (dMMSBM)***. Xing et al. (2010) developed a Bayesian state-space approach for modeling the evolution of the underlying roles of entities in a network, such that a network evolves in time through the random walk dependence structure on the hyperparameters of the prior distribution of the membership vectors and the block probabilities. Sarkar and Moore (2005) introduced a predictive latent space model for temporal networks. Similar work on extending latent space model to account for longitudinal networks is also done by Sewell and Chen (2014). However, none of the latent variable models for temporal networks focus on estimating covariate effects on which most of the questions in education research hinges.

In this paper, we focus on ***latent space network models***, and address some of the drawbacks of the existing methods to make the model applicable for networks in education research. Hoff et al. (2002) introduced latent space model for social networks, and Hoff and Ward (2004); Handcock et al. (2007); Raftery et al. (2012) have explored methodological and computational aspects of the static latent space models. We also present the utility of the model by analyzing advice seeking network of teachers (Spillane et al., 2012) observed over 5 time points.

3 Model

The ***latent space network model (LSM)*** introduced by Hoff et al. (2002) is characterized by the positions of nodes in a low-dimensional latent space. Hoff et al. (2002) describes the latent space as a social space containing the unobserved characteristics of the network, where nodes with similar latent characteristics will have nearby latent positions. Further, conditional on the latent positions the ties in a network are assumed to form independently, and the probability of a tie between nodes i and j is inversely related to the interdistance between their latent positions. LSM accounts for basic network properties like ***reciprocity, transitivity and clustering***. If two nodes share a tie, they lie close to each other in the latent space which indicates high probability of reciprocating the tie. If two ties in a network share a common node, then the two remaining nodes will lie close to each other in the latent space hence increasing the probability of a tie between those nodes. These properties are described in more detail in Hoff et al. (2002)

Let Y denote a random variable representing a network graph. We use upper-case Y to denote a random variable, and lower-case y to denote its realization. For a static network, ***an observed network*** is a graph with n vertices. These vertices are called ***the nodes*** of the network, and the corresponding edges connecting the vertices are ***the ties***. A network is usually represented by a $n \times n$ sociomatrix Y with entries Y_{ij}, where Y_{ij} measures the strength of a relationship from node i to node j, and can be either discrete or continuous. We call Y_{ij} a tie from i to j. A tie Y_{ij} and its reciprocal tie Y_{ji} collectively form ***a dyad*** between i and j. For an undirected network, $Y_{ij} = Y_{ji}$. For simplicity, we will consider a discrete network Y such that

$$Y_{ij} = \begin{cases} 1 \text{ if there is a tie from i to j} \\ 0 \text{ if there is no tie.} \end{cases}$$

However, the models discussed in the paper can be easily extended to ordinal and continuous valued ties based on the techniques used for generalized linear models. We will use Z to denote a $n \times d$ matrix of the latent positions, such that its i^{th} row Z_i is a vector representing the position of a node i in a d dimensional latent space.

We can represent the LSM model in notation in Equation 1:

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(p_{ij}) \\ \eta_{ij} &:= \text{logit}\,(p_{ij}) = \beta_0 - ||Z_i - Z_j|| \\ Z_i &\sim MVN(0, \Sigma). \end{aligned} \tag{1}$$

Here, p_{ij} denotes the probability of forming a tie from node i to j and $||Z_i - Z_j||$ denotes the distance (for example, Euclidean) between the latent positions Z_i and Z_j. Finally, β_0 is an overall intercept of the model.

We can then invoke the conditional independence of ties given latent positions and intercept to get the likelihood of the observed network y conditional on the latent positions Z and the intercept β_0, as given by Equation 2:

$$P(Y = y|Z, \beta_0) = \prod_{i \neq j} exp[\eta_{ij} y_{ij} - \log(1 + exp(\eta_{ij}))]. \quad (2)$$

The intercept β_0 in the model can be seen as an overall fixed network effect, whereas the distance between the latent positions Z_i have the dual interpretation of being latent variable and residuals in the model. Recall that the latent positions model the unobserved characteristics of the nodes in a low dimensional Euclidean space via the interdistance of the positions. As we include observed covariates in the model that explain some of the structures in the network, any extra structure of the network that is not explained by the covariates will then be accounted for by the positions of the nodes in latent space through their interdistances. LSM is arguably a useful and appealing method for network analysis because it implicitly models different network features while making fewer assumptions about the dependence structure of the ties.

We denote edge covariates by an array X of dimension $n \times n \times K$, where K is the number of covariates in the model. X_{ijk} is used to denote the k^{th} covariate for a tie where i is the sender and j is the receiver, and β^k is the corresponding slope coefficient. We can rewrite the logistic link function in Equation 1, as done previously in Sweet et al. (2013) for hierarchical network models, to include observed covarites effect as:

$$\text{logit}(p_{ij}) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ijk} - ||Z_i - Z_j||.$$

We combine ideas from the latent space model of Hoff et al. (2002) and the state-space modeling approach to model the evolution of networks in time through the changes in the latent positions. In this paper, we assume random walk plus noise evolution of the latent positions over time.

In case of a temporal network, we define Y_t as a sociomatrix of the network at time t with entries Y_{ijt} measuring a relationship from node i to j at that time t. Further, we will use $Y_{1:T} := [Y_1 Y_2 \dots Y_T]$ to denote the list of socio-matrices up to T time point, and $Z_{1:T} := [Z_1 Z_2 \dots Z_T]$ to denote the list of latent positions up to time T. Note that, under our method it is not necessary to have same number of nodes at all T time points.

LSM for the static model can be extended to account for the temporal network with changing covariates effect in Equation 3 :

$$
\begin{aligned}
y_{ijt} &\sim Bernoulli(p_{ijt}) \text{ for } i \neq j \\
logit(p_{ijt}) &= \beta_0 + \sum_{k=1}^{K} X_{ijt}^k \beta_t^k - ||Z_{it} - Z_{jt}|| \\
Z_{i,1} &\sim MVN(0, \Sigma), \text{ for } i = 1, \dots, n \\
Z_{i,t} &= Z_{i,t-1} + \epsilon_t \text{ for } t = 2, \dots, T \\
\epsilon_t &\sim MVN(0, \Sigma).
\end{aligned} \tag{3}
$$

Here, K is the total number of edge covariates in the model, X_{ijt}^k represents k^{th} covariate related to the tie from i to j at time t and β_t^k is the corresponding slope coefficient. In the current application using advice seeking network of teachers, we are interested in seeing how β_t^k changes with t for each co-variate k.

4 Estimation

We use Metropolis Hastings within Gibbs algorithm to draw samples from the posterior distribution of the parameters, namely, the slope coefficients $\beta_{1:T}$, the intercept β_0, the latent positions up to time T, $Z_{1:T}$, and the variance covariance matrix of the latent positions, Σ. Using the conditional independence property of the network ties we can write conditional likelihood of the data as

$$\mathsf{P}(Y|Z_1,\ldots,Z_T,X,\beta,\beta_0) \;=\; \prod_{t=1}^{T}\prod_{i\neq j}\mathsf{P}(Y_{i,j,t}|Z_{i,t},Z_{j,t},X_{ijt},\beta_t,\beta_0).$$

The Markovian property of the trajectory of the latent positions allows use to write the joint density of the latent positions upto time T as

$$\mathsf{P}(Z_1,\ldots,Z_T) \;=\; \prod_{i=1}^{n}\mathsf{P}(Z_{i,1}|\mu_z,\Sigma)\times\prod_{t=2}^{T}\prod_{i=1}^{n}\mathsf{P}(Z_{i,t}|Z_{i,t-1},\Sigma),$$

where, μ_z is the mean and Σ is a diagonal variance-covariance matrix of the multivariate normal distribution of the latent positions at time 1.

We then have the full likelihood of the data and the parameters into a simple product form as in Equation 4.

$$\begin{aligned}\mathsf{P}(Y,Z,X,\beta,\beta_0) &= \mathsf{P}(Y|Z_1,\ldots,Z_T,X,\beta,\beta_0)\mathsf{P}(Z_1,\ldots,Z_T|\beta_0)\mathsf{P}(\beta_0)\\ &= \prod_{t=1}^{T}\prod_{i\neq j}\mathsf{P}(Y_{i,j,t}|Z_{i,t},Z_{j,t},X_{ijt},\beta_t,\beta_0)\times\prod_{i=1}^{n}\mathsf{P}(Z_{i,1}|\mu_z,\Sigma)\\ &\times\prod_{t=2}^{T}\prod_{i=1}^{n}\mathsf{P}(Z_{i,t}|Z_{i,t-1},\Sigma)\times\mathsf{P}(\beta_0)\times\mathsf{P}(\Sigma).\end{aligned} \tag{4}$$

Following prior distributions are specified to draw samples from the posterior distribution of the parameters given data using MCMC algorithm:

$$
\begin{aligned}
Z_{i1} &\sim \text{Multivariate Normal}(\mu_z, \Sigma) \\
\beta_0 &\sim \text{Normal}(\mu_0, \sigma_0^2) \\
\beta_t^k &\sim \text{Normal}(\mu_0, \sigma_0^2) \forall k, t \\
\Sigma_{ii} &\sim \text{InverseGamma}(A, B).
\end{aligned}
$$

μ_0,σ_0^2 are prior mean and variance of the intercept and slope parameters, whereas μ_z and Σ are hyperparameters of the model specifying mean and variance-covariance matrix of the prior distribution of the latent positions at time $t = 1$. A and B are the shape and the rate parameters of the Inverse-Gamma prior distribution on Σ. We constrain Σ to be a diagonal matrix to simplify computation.

Note that, since the likelihood of the data is related to the latent positions only through the latent distance, the latent positions Z_t's are identifiable only upto a distance preserving transformation. For inference and summarizing the MCMC posterior draws of the latent positions, we address the issue of unidentifiability using procrustes transformation (Borg and Groenen, 2005) of the posterior draws as a post-processing step. Following Hoff et al. (2002), we initialize the latent positions in the MCMC using multidimensional scaling on the observed distance matrix (minimum path length between nodes). We also use these positions at each time t as a target for procrustean transformation. Similar approach has been used by Hoff et al. (2002) and Sewell and Chen (2014).

MCMC algorithm for LLSM is coded in *Rcpp* (Eddelbuettel and François, 2011) and *R* (R Core Team, 2014), and is available upon request.

4.1 Estimation with Missing Nodes

The first term of the product in Equation 4 can be easily obtained from Equation 2 for each time point t since the networks are independent over time conditional on the covariates, the latent positions and the intercept. However, we need to account for the changing number of the nodes over time while computing the likelihood of the latent positions, which is a major concern in many real world applications. We will assume that nodes are missing at random. Further, once a node exits a network there is a very less chance that it will re-enter. Thus, if a node enters the network at $t > 1$, we will use the t at which the node enters the network as its initial time and assume that its latent position has the same prior as Z_{i1}. Let $\{N_{t-1}\}$ denote set of nodes at time t that were also present at $t-1$.

The likelihood in Equation 4 can be re-written as:

$$\begin{aligned}\mathsf{P}(Y, Z, X, \beta, \beta_0) = &\prod_{t=1}^{T}\prod_{i\neq j}\mathsf{P}(Y_{ijt}|X_{ijt}, z_{it}, z_{jt}, \beta_t, \beta_0) \times \prod_{i=1}^{n_1}\mathsf{P}(z_{i,1}|\mu_z, \Sigma)\\ &\times \prod_{t=2}^{T}\prod_{i=1}^{n_t}[\mathsf{P}(z_{i,t}|z_{i,t-1}, \Sigma)\mathsf{I}(i \in \{N_{t-1}\}) + \mathsf{P}(z_{i,t}|\mu_z, \Sigma)\mathsf{I}(i \in \{N_{t-1}\})]\\ &\times \mathsf{P}(\beta_0) \times \mathsf{P}(\beta) \times \mathsf{P}(\Sigma)\end{aligned} \tag{5}$$

In this setup, we do not necessarily require the nodes to be present at all time points. Assuming that some of the nodes are missing at random, we will use the time point at which a node appear in the network as its initial time and hence the latent positions for this node at the initial time point are centered around μ_z. Bayesian framework of our model also allows for easy imputation of the missing ties.

5 Data analysis

To illustrate the utility of our model in education research, we use advice-seeking network data of teachers in primary schools from a school district (Spillane et al., 2012),(Spillane et al., 2012),(Spillane and Hopkins, 2013). Following the convention in Spillane et al. (2012) and Sweet (2015), we will use the pseudonym Auburn Park as the name of the district. Spillane et al. (2012) and Sweet (2015) have analyzed this network of teachers at a particular time point, concluding that teachers teaching in the same grade and of similar gender are more likely to interact with each other for advice. However, longitudinal analysis of the advice-seeking networks is missing. Number of changes were made in the schools since the first year of data collection. Assessment was changed from district level to state level, first for Reading and then for Math in the following year; new math coaches were introduced in the schools after the first year of data collection. We expect these changes to change teachers' incentives for communication and also their expectations from each other in terms of instructional advice. In this section, we go a step further than Spillane et al. (2012) and Sweet (2015) to investigate whether grade and gender of the teachers along with their age and years of experiences affect the likelihood of forming advice seeking ties and how this relationship changes over time.

The data includes school staff surveys as well as social network data from 14 elementary schools for 5 different academic years observed once every year. Staff members were asked to name the individuals to whom they seek instructional advice and information, and were able to nominate any other staff member in the district. Network tie density at all 5 time points are reported in Table 1. We see that the network size (as represented by its nodes) is fairly large and consistent over time, whereas the ties in the network are sparse as represented by very small tie density. Tie density is lowest at the last time point.

This duplicates material in the introduction... Maybe a better idea would be to discuss the figure from your talk that shows all the changes over 2-3 years in the district.

Year	Number of schools	Total nodes	Tie Density
2010	14	326	0.0177
2011	14	385	0.0157
2012	14	372	0.0152
2013	14	387	0.013
2015	14	387	0.010

Table 1: Network Summary

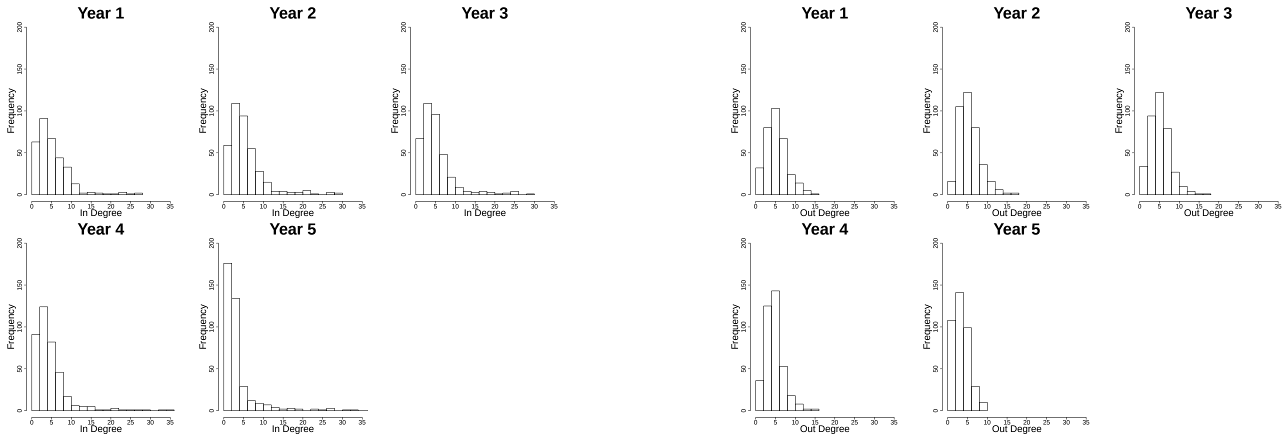
Histograms of indegree and outdegree of the nodes in the networks at each time point are displayed in Figure 2. In the figure, we can notice that the the network is asymmetric and the distribution of indegree and outdegree is very different towards the tail. This feature of the plots can be explained by the asymmetric advice seeking behavior where a person who seeks advice is not necessarily also sought after and vice-versa. In the context of these networks, the senders are the individuals seeking advice or information and the receivers are the individuals providing advice or information. The tie then indicates an advice/information relationship from the seeker to the provider.

In addition to the information about who goes to whom for advice, we also have information on the covariates of the teachers in the networks. For example, we have data on the schools the teachers teach, their gender, the grade/grades they teach, their age, and their years of experience. Nodal covariates are converted into edge covariates so that we can measure the effect of similarity and/ or differences in the nodes on the basis of their covariates in tie formation. Gender and school id of the teachers are converted into edge covariates by converting them into binary variable such that:

$$S_{ijt} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in same school at time } t \\ 0 & \text{otherwise} \end{cases}$$

and,

Here and below, more space before and after each equation is needed.



(a) (b)

Figure 2: Histograms representing the in-degree and out-degree of the nodes in the network at 5 different years. a) is the plot of in-degree at each year the data was collected, and b) is the plot of out-degree.

$$Ge_{ijt} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ have same gender at time } t \\ 0 & \text{otherwise.} \end{cases}$$

More space around equations pls!

To account for the age differences of sender and receiver teachers, a new edge covariate $AgeDiff$ is created such that:

$$AgeDiff_{ijt} = Age_{it} - Age_{jt},$$

No space here.

where Age_{it} is the age of the sender node i at time t and Age_{jt} is the age of the receiver node j at time t.

Next, to study the effect of grades on advice seeking behavior of the teachers, edge covariates Gr^l are created, for grades first to sixth, such that

$$Gr^l_{ijt} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ both teach grade } l \text{ at time } t \\ 0 & \text{otherwise.} \end{cases}$$

Here, teachers teaching prek and kindergraden are used as a base case. Not sure what you mean here.

In addition, we also have data on the years of experience of the teachers in the school (self

pre-k

reported by the teachers). To account for assymetry in the advice seeking relationship and test whether new teachers send more advice seeking tie and whether experienced teachers receive more advice seeking tie, this nodal feature is used as a sender and receiver covariate for years of experience of the sender and the receiver separately. Let, ES_{ijt} denote the experience of node i at time t where i sends advice seeking tie to node j. Similarly, let ER_{ijt} denote the experience of node j at time t, where j receives advice seeking tie from node i.

Finally, some of the covariates are missing for some nodes, either at all the time points or in some particular time point. Missingness for covariates is generally less than 5%. We have two methods of completing such missing data. First, for a node with a covariate missing at a time point, temporal property of the data is used to borrow missing information from neighboring time points for the node, if it is observed at those time points. If the covariate is missing for the node at all time points, it is replaced by the average value for the school the teacher teaches at that time point. From our previous analysis, we have observed that the teachers' advice-seeking network cluster by their school membership [see supplement materials], which led us to believe that our method is a valid approach to use network structure for missing covariate imputation. More elaborate imputation of missing covariates is a part of our future work.

5.1 Fitting LLSM to Auburn Park Data

We fit LLSM introduced in Section 3 with covariates using the model in Equation 6.

$$
\begin{aligned}
Y_{ijt} &\sim Bernoulli(p_{ijt}) \text{ for } i \neq j \\
logit(p_{ijt}) &= \beta_0 + \beta_{S(t)} S_{ijt} + \sum_{l=1}^{6} \beta^l_{Gr(t)} Gr^l_{ijt} + \beta_{Ge(t)} Ge_{ijt} + \\
&\quad \beta_{AgeDiff(t)} AgeDiff_{ijt} + \beta_{ES} ES_{ijt} + \beta_{ER} ER_{ijt} - ||Z_{it} - Z_{jt}|| \qquad (6) \\
Z_{i,1} &\sim MVN(0, \Sigma), \text{ for } i = 1, \dots, n \\
Z_{i,t} &= Z_{i,t-1} + \epsilon_t \text{ for } t = 2, \dots, T \\
\epsilon_t &\sim MVN(0, \Sigma).
\end{aligned}
$$

Parameters for the prior distributions in the model are specified as following:

$$
\begin{aligned}
Z_{i1} &\sim \text{Normal}(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma) \\
\beta_0 &\sim \text{Normal}(0, 1000) \\
\Sigma_{ii} &\sim \text{InverseGamma}(100, 150) \\
\beta^k_t &\sim \text{Normal}(0, 1000) \forall k, t
\end{aligned}
$$

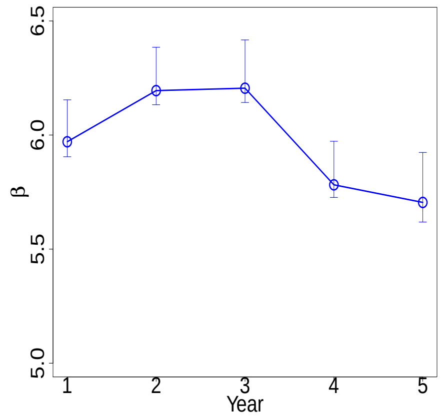
We use Metropolis Hastings within Gibbs Algorithm to sample draws from the posterior distribution of the parameters given the data. The Metropolis-Hastings draws were tuned to ensure acceptance rates between 0.38 and 0.46 for one-dimensional draws and between 0.2 and 0.28 for drawing parameters with higher dimensions follwing Gelman et al. (2003). We generated 7000 MCMC draws to ensure convergence. After burnin of the first 2000 draws and thinin of 20, we retained a total of 250 sample. Posterior mean of the MCMC draws of the coefficients along with 95% (equaltailed) credible interval for each covariate over five different time points are displayed in Figure 3.

Posterior mean of the slope parameters along with their 95% credible interval displayed in Figure 3 suggest that, teaching in the same school is a very strong indicator of whether there exists an advice seeking tie between two teachers. However there is a drop in the effect of the school covariate after time point 3. Also teachers are more likely to go for advice to someone of same gender or teaching in similar grade. However, the effect of teaching in same grade seems to been different for different grades over time. Further, while the effect size is small teachers are more likely to go to someone who is older than them for advice, as shown by the negative and non-zero coeffcient for covariate defined by the difference in age of the sender and the receiver. This effect is also mostly consistent over time. Experience of the sender and receiver nodes seem to have interesting and unexpected effect. In the first two time points, teachers with fewer years of experiences are more likely to send advice seeking ties and the teachers with more years of experiences are more likely to receiver advice seeking ties. However, the relationship is insignificant or reversed in later years of the survey.

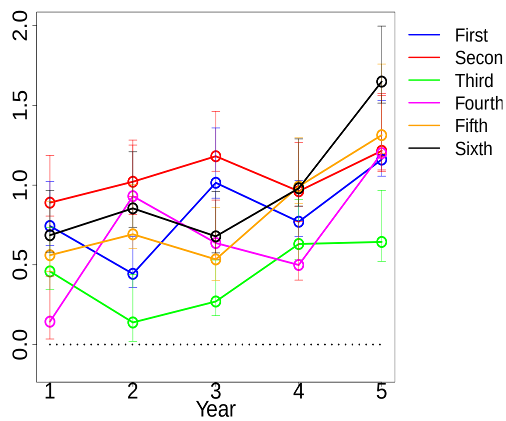
Next we also look at the estimated latent positions that explains latent structure of the networks not explained by the observed covariates in the model. Posterior mean along with 1 posterior standard error ellipse of the latent positions over 5 years are displayed in Figure 4. We observe that the spread of the estimated latent positions increases over time, which might explain the decrease in tie density of the advice seeking network over time.

5.2 Posterior Predictive Checking

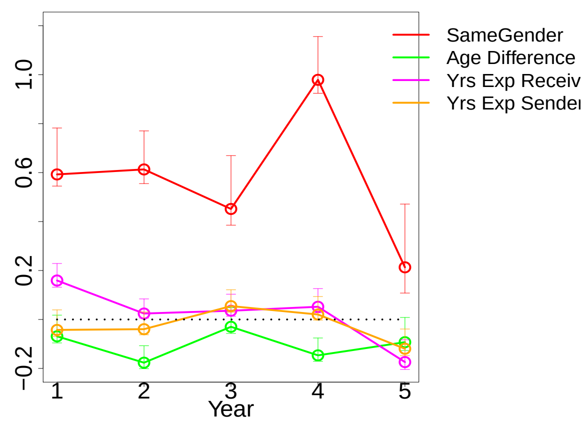
We use posterior predictive assessment following the guidelines in Gelman (2003) to compare the fits from the MCMC algorithm and the observed data. Let $\hat{p} = p(yrep|y)$ denote



(a) Same School



(b) Grades



(c) Gender, Age & Years of Experiences

Figure 3: Posterior mean of the coefficients along with 95% credible interval from LLSM for different edge covariates; a.) effect of teaching in same school vs. different schools over 5 different years, b.) effect of the interaction between grade level and whether two teachers teach in same grade over 5 years and c.) effects of similarity in the gender, age difference between sender and receiver of the ties and the years of experiences of the sender and receiver separately, over the 5 different year.
Y-axis is the estimated value of the β coefficient, with scale adjusted to account for different ranges of the set of covariates shown in each plot.

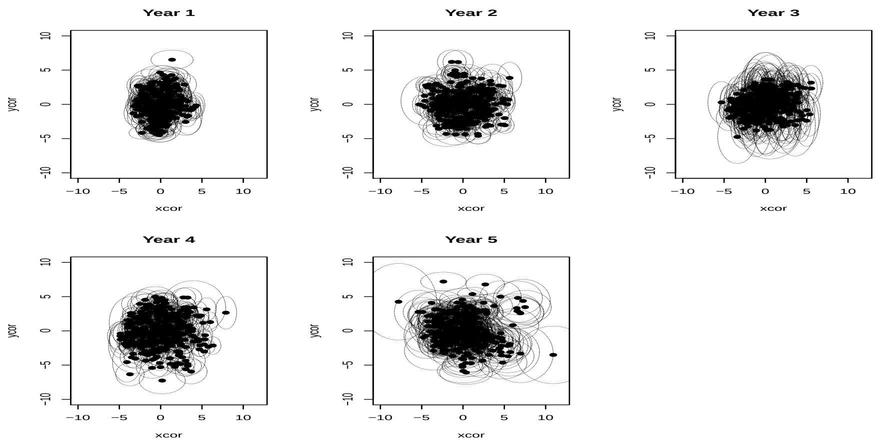


Figure 4: Posterior mean along with 1 posterior standard error ellipse of the latent positions over 5 years.

distribution of future data conditional on the observed data under the model being fit. This is the posterior predictive distribution of Y. Then we can assess the fit of the model by comparing $\hat{p}$ with observed network data at each time point.

Let $\theta_t = \{Z_t, \beta_0, \beta_t\}$ denote the set of model paramters in LLSM at time t. Then we can compute $\hat{p}_t$ for each time point t as in Equation 7.

$$\begin{aligned}\hat{p}_t &= \mathsf{P}(y_t^{\text{rep}}|Y) \\ &= \int \mathsf{P}(y_t^{\text{rep}}|\theta_t)\mathsf{P}(\theta_t|Y)d\theta_t \\ &\approx \frac{1}{L}\sum_{l=1}^{L}\mathsf{P}(y_t^{\text{rep}}|\theta_t^{(l)}).\end{aligned} \tag{7}$$

Here, L is the total number of MCMC draws from the posterior distribution of the paramters given data.

In figure 5, we compared the observed data with the posterior predictive probabilities of ties in each year using the image plots. The ties are plotted by the school membership of the teachers to show a clear block structure in the data by school. We plotted the predicted probabilities in the same order as well. It is clear from the figure that the model recovers the observed data very well.

Would be useful to compare these plots with similar plots for a model that clearly doesn't fit.

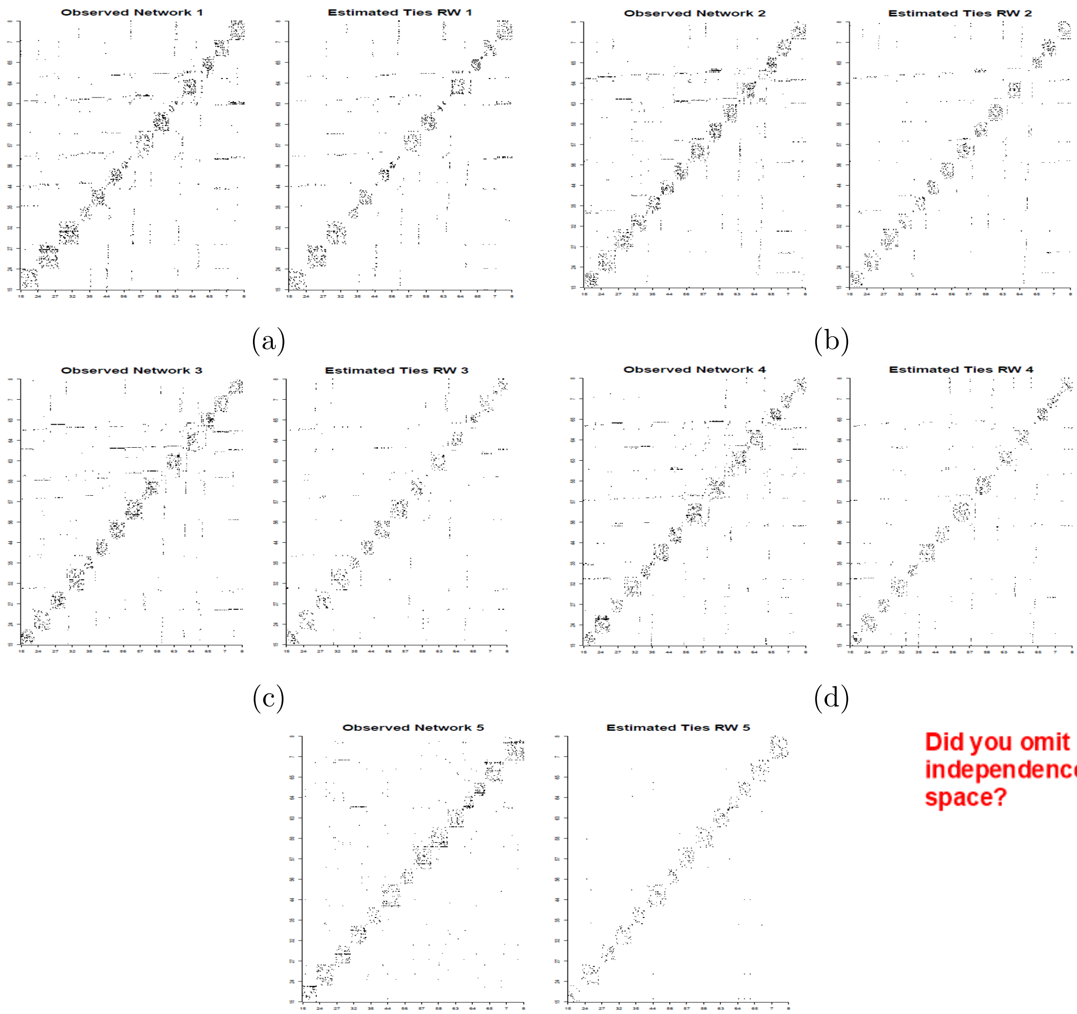
5.3 Comparision with Model Fits from Models with Assumption of Independences of Ties

For completeness of our analysis, we compared the fits from simpler models to the Auburn Park data with the fits from our model. First we fitted time independent logistic regression with covariates at each time point (refer to Equation 8) and computed fitted probability using the point estimates of the model fits.

$$logit(P(Y_{ijt} = 1)) = \beta_0 + \beta_{S(t)} S_{ijt} + \sum_{l=1}^{6} \beta^l_{Gr(t)} Gr^l_{ijt} + \beta_{Ge(t)} Ge_{ijt} + \beta_{AgeDiff(t)} AgeDiff_{ijt} + \beta_{ES} ES_{ijt} + \beta_{ER} ER_{ijt} \tag{8}$$

Second, we fitted logistic regression using prior ties information as a covariate as shown in Equation 9. This is a modified autoregressive logistic model. Again, we used the point estimates of the coefficients at each time point to compute fitted probabilities of tie.

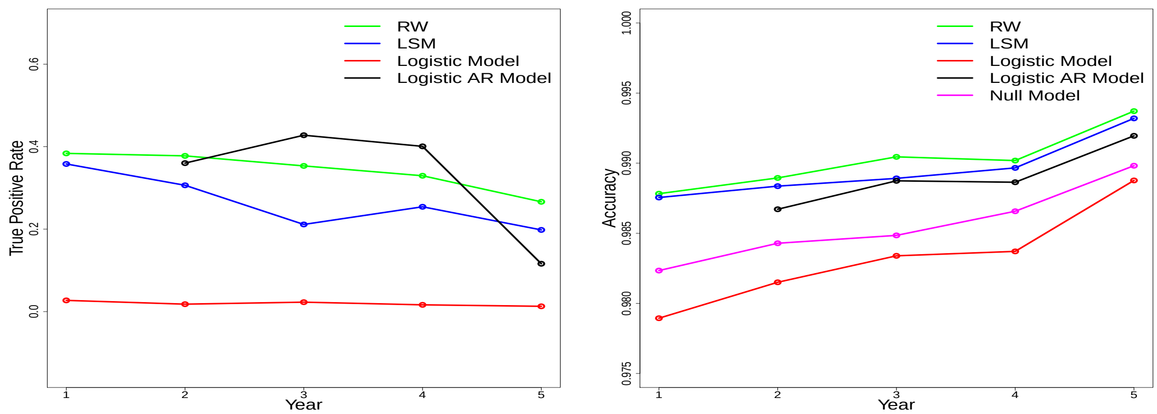
$$logit(P(Y_{ijt} = 1)) = \beta_0 + \beta_{t-1} Y_{ij(t-1)} + \beta_{S(t)} S_{ijt} + \sum_{l=1}^{6} \beta^l_{Gr(t)} Gr^l_{ijt} + \beta_{Ge(t)} Ge_{ijt} + \beta_{AgeDiff(t)} AgeDiff_{ijt} + \beta_{ES} ES_{ijt} + \beta_{ER} ER_{ijt} \tag{9}$$



(a) (b) (c) (d) (e)

Did you omit plots for simpler independence models to save space?

Figure 5: Image plots for comparing observed network and estimated ties using the posterior predictive fits from LLSM at a.) Year 1, b.) Year 2, c.) Year 3, d.) Year 4 and e.) Year 5



(a) True Positive Rates

(b) Accuracy

Figure 6: Plots Comparing a.) true positive rates and b.) Accuracy, of the estimates from different models used to fit advice seeking networks. Posterior predictive distribution is used to computed fitted probabilties of ties for the two Bayesian methods, LSM and LLSM RW.

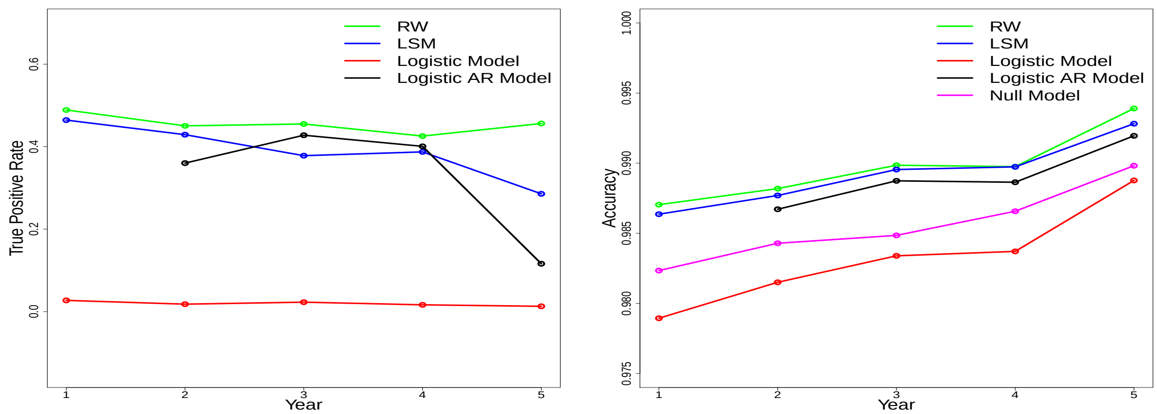
Finally, we fitted independent latent space model for each time point without the random walk dependence of the latent positions using the model shown in Equation 1.

For independent LSMs and our proposed longitudinal LSM, we used estimated probabilities of ties computed from the posterior predictive distribution. We used accuracy or the percentage of correctly classified presence and absence of ties in observed data, and true positive rate or the proportion of correctly classified observed ties as two metrics for model evaluation, which are displayed in Figure 6.

In-sample fits from point estimates are often known to be optimistic [[TODO:CITE]]. So, to be fair on our Bayesian estimates we also compared fitted probabilities using posterior mean as the point estimates for LSM and LLSM. Corresponding measure of accuracy and true positive rates using point estimates are shown in Figure 7

Model assesessments as displayed in Figures 6 and 7 show how thet overall fit improves

I'm not sure I agree that plugging in point estimates Corrects in-sample bias.



(a) True Positive Rates

(b) Accuracy

Figure 7: Plots Comparing a.) true positive rates and b.) Accuracy, of the estimates from different models used to fit advice seeking networks. Point estimates of the parameters are used to compute fitted probabilities of ties for all the models under consideration.

by using the latent space model and by accounting for the time dependence.

6 Conclusion

References

Airoldi, E. M., D. M. Blei, S. E. Fienberg, E. P. Xing, and T. Jaakkola (2006). Mixed membership stochastic block models for relational data with application to protein-protein interactions. In *In Proceedings of the International Biometrics Society Annual Meeting*.

Borg, I. and P. J. Groenen (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer.

Dabbs, B., S. Adhikari, M. Sadinle, B. W. Junker, T. M. Sweet, and A. C. Thomas (in prep.). Conditionally independent network models.

Eddelbuettel, D. and R. François (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software 40*(8), 1–18.

Fienberg, S. E. (2012, October). A Brief History of Statistical Models for Network Analysis and Open Challenges. *Journal of Computational and Graphical Statistics 21*(4), 825–839.

Frank, O. and D. Strauss (1986, September). Markov graphs. *Journal of the American Statistical Association 81*(395), 832–842.

Gelman, A. (2003). A bayesian formulation of exploratory data analysis and goodness-of-fit testing*. *International Statistical Review 71*(2), 369–382.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003, July). *Bayesian Data Analysis, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)* (2 ed.). Chapman and Hall/CRC.

Gill, P. S. and T. B. Swartz (2001). Statistical analyses for round robin interaction data.

The Canadian Journal of Statistics / La Revue Canadienne de Statistique 29(2), pp. 321–331.

Goldenberg, A., A. X. Zheng, S. E. Fienberg, and E. M. Airoldi (2010, feb). A survey of statistical network models. *Found. Trends Mach. Learn. 2*(2), 129–233.

Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 170*(2), 301–354.

Handcock, M. S., G. Robins, T. A. Snijders, J. Moody, and J. Besag (2003). Assessing degeneracy in statistical models of social networks. Technical report, Citeseer.

Hanneke, S., W. Fu, E. P. Xing, et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics 4*, 585–605.

Hanneke, S. and E. P. Xing (2007). Discrete temporal models of social networks. In *Proceedings of the 2006 Conference on Statistical Network Analysis*, ICML'06, Berlin, Heidelberg, pp. 115–125. Springer-Verlag.

Harris, K. M., C. T. Halpern, B. C. Haberstick, and A. Smolen (2013, 2). The national longitudinal study of adolescent health (add health) sibling pairs data. *Twin Research and Human Genetics 16*, 391–398.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 97*, 1090–1098.

Hoff, P. D. and M. D. Ward (2004). Modeling dependencies in international relations networks. *Political Analysis 12*(2), 160–175.

Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association 76*(373), pp. 33–50.

O'Malley, J. A. (2013). The analysis of social network data: as exciting frontier for statisticians. *Statistics in Medicine 32*(4), 539–555.

Penuel, W. R., M. Riel, A. E. Krause, and K. A. Frank (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *The Teachers College Record 111*, 124163.

Pitts, V. M. and J. P. Spillane (2009). Using social network methods to study school leadership. *International Journal of Research & Method in Education 32*(2), 185–207.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A. E., X. Niu, P. D. Hoff, and K. Y. Yeung (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics 21*(4), 901–919.

Robins, G. and P. Pattison (2001). Random graph models for temporal processes in social networks. *The Journal of Mathematical Sociology 25*(1), 5–41.

Robins, G., P. Pattison, Y. Kalish, and D. Lusher (2007, May). An introduction to exponential random graph (p*) models for social networks. *Social Networks 29*(2), 173–191.

Sarkar, P. and A. W. Moore (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD EXPLORATIONS NEWSLETTER 7*(2), 31–40.

Sewell, D. K. and Y. Chen (2014). Latent space models for dynamic networks. *Journal of the American Statistical Association* (ja), 00–00.

Snijders, T. A. (1996). Stochastic actor-oriented models for network change. *Journal of mathematical sociology 21*(1-2), 149–172.

Spillane, J. P. and M. Hopkins (2013). Organizing for instruction in education systems and school organizations: How the subject matters. *Journal of Curriculum Studies 45*(6), 721–747.

Spillane, J. P., C. M. Kim, and K. A. Frank (2012). Instructional advice and information providing and receiving behavior in elementary schools exploring tie formation as a building block in social capital development. *American Educational Research Journal*.

Spillane, J. P., P. Peterson, M. Sherin, S. Fisher, and S. Konstantopoulos (2012). Reta: Chicago school staff social network questionnaire longitudinal study, 2005-2008.

Sweet, T. M. (2015). Incorporating covariates into stochastic blockmodels. *Journal of Educational and Behavioral Statistics*, 1076998615606110.

Sweet, T. M., A. C. Thomas, and B. W. Junker. Hierarchical mixed membership stochastic blockmodels for multiple networks and experimental interventions. *Handbook of Mixed Membership Models*.

Sweet, T. M., A. C. Thomas, and B. W. Junker (2013). Hierarchical network models for education research: Hierrarchical latent space models. *Journal of Educational and Behavioral Statistics 38*(3), 295–318.

Westveld, A. H. and P. D. Hoff (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics 5*(2A), 843–872.

Xing, E. P., W. Fu, and L. Song (2010, 06). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics 4*(2), 535–566.