POWER TO DETECT INTERVENTION EFFECTS ON ENSEMBLES OF SOCIAL NETWORKS

Tracy M. Sweet and Brian W. Junker

September 24, 2013

Correspondence should be sent to Tracy M. Sweet

E-Mail: tsweet@umd.edu
Phone: 301-405-3623
Fax: 301-314-9245
Website: http://education.umd.edu/EDMS/fac/Sweet/

This work was supported by Program for Interdisciplinary Education Research Grant R305B040063 and Hierarchical Network Models for Education Research Grant R305D120004, Institute for Education Sciences, Department of Education.

POWER TO DETECT INTERVENTION EFFECTS ON ENSEMBLES OF SOCIAL NETWORKS

Abstract

The hierarchical network model (HNM) is a framework introduced by (Sweet, Thomas & Junker, 2013) and (Sweet, Thomas & Junker, 2014) for modeling interventions and other covariate effects on ensembles of social networks, such as what would be found in randomized controlled trials in education research. In this paper, we develop calculations for the power to detect an intervention effect using the hierarchical latent space model (HLSM), an important subfamily of HNMs. We derive basic convergence results and asymptotic bounds on power, showing that standard error for the treatment effect is inversely proportional to the product of the number of ties and the number of networks; a result rather different from the usual effect of cluster size in hierarchical linear models, for example. We explore these results with a simulation study and suggest a tentative approach to power for practical applications.

Key words: asymptotics, bayesian statistics, experimental design, hierarchical latent space model, hierarchical network model, power, sample size, social network analysis

I. Introduction

Many large-scale educational interventions are applied to entire schools or entire grades within a school (e.g., Matsumura, Garnier & Resnick, 2010; Spillane, Correnti & Junker, 2009; Hord, Roussin & Sommers, 2010; Glennan & Resnick, 2004; McLaughlin & Talbert, 2006) and target changes in the organizational structure or professional climate in the school. Social networks are an obvious outcome measure for interventions aimed at changing the social structure in the school since the network of relationships among teachers, administrators and other school staff not only reveals how information or resources are shared but also the overall professional structure of the school: whether teachers are well-connected, isolated into subgroups, completely isolated, etc. For example, an intervention whose aim is to improve professional working relationships among school staff can be assessed through changes to the advice-seeking networks (Spillane, Correnti & Junker, 2009). Other professional relationship ties include collaboration and co-teaching. Interventions may also target student populations. An intervention aimed at reducing bullying and sexual aggression among adolescents (e.g. Espelage & Low, 2009) is likely to change the way students interact with each other, increasing the frequency of positive interactions and friendships.

Sweet et al. (2013, 2014) introduce a modeling framework for estimating treatment effects for experiments on networks as well as a statistical network model that can be used for both observational and intervention network data, and Sweet, Thomas & Junker (2012) introduce a model that can capture intervention effects on network subgroup structure.

However, virtually no work has been done to incorporate networks into experimental design. Given that the network is the choice outcome of an organizational experiment, the experiment should be powered so that intervention effects are detectable on the network.

The purpose of this paper is to initiate a body of research on experimental design for social networks, beginning with power analysis methodology for sample size calculations. In Section 2, we describe the work done by Sweet et al. (2013) and present a simple model for interventions on networks, the hierarchical latent space model (HLSM) and explain why power calculations used for hierarchical linear models are not applicable for social networks. We present theoretical results in Section 3. Then in Section 4, we present power calculations for one type of network model and assess our analytical results using a series of simulation studies presented in Section 6. Finally, we conclude by discussing practical applications and future work.

We present a series of results. We show asymptotic consistency and normality for treatment effects in multiple network intervention studies. We also derive bounds on treatment effect standard error and power and we show that sample size of each network, i.e. the number of nodes in each network, is much more influential in determining power than cluster size in an HLM.

2. Hierarchical Network Models

Sweet et al. (2013) introduced a framework called Hierarchical Network Models (HNMs) to build multilevel models for ensembles of social networks for network level experiments or observational studies. Let $\mathbb{Y}$ be the collection of observed networks $(Y_1, \ldots Y_K)$ and let $\mathbb{X}$ be the collection of factors associated with observing a tie. For example, Y_k may be a friendship network in school k. Y_{ijk} represents the relationship or tie from person i to person j in that school. Likewise, $\mathbb{X}$ may be individual-, tie- or network-level covariates believed to be associated with friendship, such as age, being the same gender, or school parochial status.

We write a HNM as follows:

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

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

where $P(Y_k|X_k, \Theta_k)$ is a model for a single network Y_k, $\Theta = (\theta_1, \ldots, \theta_k)$ is the collection of model parameters that come from some distribution F, and $W_1, \ldots, W_K$ are covariates. In general, F is quite flexible in that it can include covariates W_k to model dependence assumptions across the networks and/or include parameters ψ to accommodate additional hierarchical structure.

To model an intervention, we include a treatment effect parameter α and treatment group covariate T_k and resulting model can be written as

$$P(\mathbb{Y}|\mathbb{X},\Theta) = \prod_{k=1}^{K} P(Y_k|X_k = (X_{1k},\ldots,X_{P_k}),\Theta_k = (\theta_{1k},\ldots,\theta_{Q_k},\alpha))$$

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

2.1 Hierarchical Latent Space Models

The HNM framework allows users a variety of choice of models for estimating intervention effects. For the purposes of this paper, we restrict our methods to a simple version of a single model, the hierarchical latent space model (HLSM; Sweet et al., 2013).

The HLSM is a multilevel extension of latent space models introduced by Hoff, Raftery & Handcock (2002). There are two main assumptions of these models: first each individual in the network occupies a position in a latent social space and the distance between any two individuals' positions influences the probability of a tie between them. The second is that ties are independent conditional on the latent space positions.

A simple HLSM for an intervention is given as

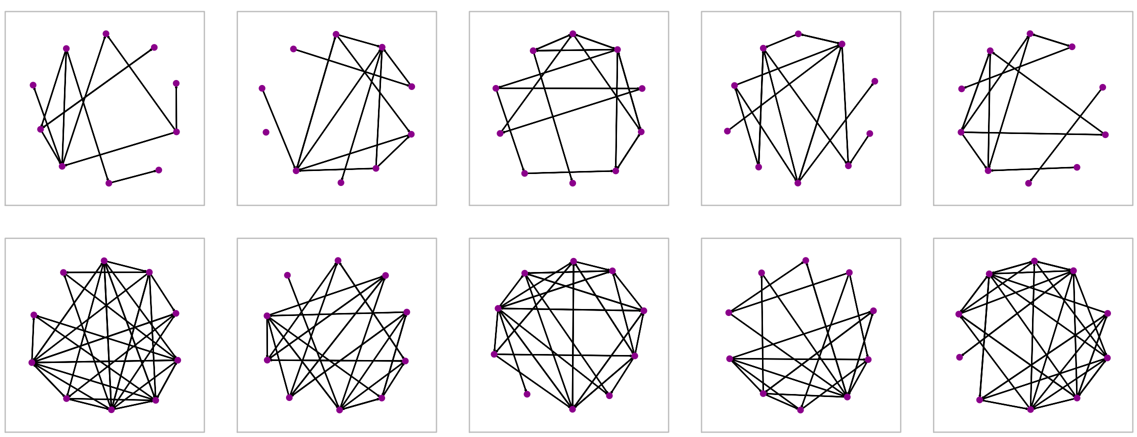
$$\begin{aligned} P(\mathbb{Y}|\beta_0,\beta_1,\theta,\mathbb{Z}) &= \prod_k \prod_{i\neq j} P(Y_{ijk}|Z_{ik},Z_{jk},\beta_0,\beta_1,\theta) \\ \text{logit}\, P[Y_{ijk}=1|\beta_0,\beta_1,\theta,Z_k] &= \beta_0 + \beta_{1k}W_{ijk} + \theta X_k - |Z_{ik}-Z_{jk}| \\ \beta_p &\sim N(\mu_{\beta_p},\tau_{\beta_p}) \\ \theta &\sim N(\mu_\theta,\tau_\theta) \\ Z_{ik} &\sim MVN(\mu_Z,\Sigma_Z) \ , \end{aligned} \tag{1}$$

where Z_{ik} and Z_{jk} are the latent space positions for individuals i and j respectively in network k, β_0 is an overall density term, β_1 is a set of regression coefficients for possibly vector-valued covariates W, and θ is the treatment effect. If we assume two conditions, then $X_k \in \{0, 1\}$ is an indicator for being in one condition, and the treatment affects the overall probability of a tie. Note also that parameters may have alternative prior distributions.

Figure 1 displays networks generated from the HLSM given by logit $P[Y_{ijk} = 1] = 1.5 + 0.5X_k - |Z_{ik} - Z_{jk}|$ where $X_k = 1$ denotes the treatment condition. Since the treatment effect is positive, we expect an increase in overall probability of a tie and in fact, the networks in the treated condition (bottom row) are visually more dense (i.e. have more ties) than the networks in the control condition (top row).

Figure 1.

Networks simulated from the HLSM given by logit $P[Y_{ijk} = 1] = 1 + 0.75X_k - |Z_{ik} - Z_{jk}|$ where $X_k = 1$ denotes the treatment condition. The treatment effect increases the probability of a tie and networks in the treatment condition (bottom row) have more ties than networks in the control condition (top row).



Power Calculations for HLMs Are Not Applicable to HNMs

The power function for a hypothesis test ϕ is defined as $\beta_\phi(\theta) = E_\theta[\phi(X)]$ where X is the observed data and $\theta \in \Theta$ is the parameter in the hypothesis (Schervish, 1997). Generally, we consider the hypothesis test ϕ as having some rejection region R and test statistic $T(X)$ and define the power function as $\beta(\theta) = P(T(X) \in R|\theta)$ (Casella & Berger, 2002).

For example, the level-α Wald test contrasting the hypotheses

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$

has a test statistic $T(X) = \hat{\theta}$ and rejection region $R = \{\hat{\theta} : \hat{\theta} > \theta_0 + Z_\alpha SE(\hat{\theta})\}$ for a suitable cutoff Z_α. The power function is then $\beta(\theta) = P\left(\hat{\theta} > \theta_0 + Z_\alpha SE(\hat{\theta})|\theta\right)$. The level of the test is $\beta(\theta_0)$, the probability of rejecting the null when the null is true. Thus, $SE(\hat{\theta})$ inversely influences the power function.

To contrast power calculations for HNMs with HLMs, consider a simple 2-level randomized cluster trial with T clusters and s individuals per cluster,

$$Y_{ij} = \theta_{0j} + e_{ij} \;\; i = 1, \ldots, s, \;\; j = 1, \ldots, T$$
$$\theta_{0j} = \gamma + \theta W_j + u_{0j},$$

where θ is the treatment effect and $e_{ij} \sim N(0, \sigma^2)$ and $u_{0j} \sim N(0, \tau^2)$. Then $SE(\hat{\theta})$ can easily be calculated and in fact,

$$SE(\hat{\theta})^2 = \frac{4(\tau + \frac{\sigma^2}{s})}{T} \; . \tag{2}$$

We use this example to highlight two important points. First, this calculation employs the HLM assumption that within cluster residuals are independent. This assumption is unrealistic for HNMs because within cluster (network) residuals are structured by the presence or absence of ties, which are rarely independent. For example, a friendship tie between A and B and between B and C is likely to affect the probability of a tie between A and C. Many current social network models accommodate dependence in a variety of ways (Frank & Strauss, 1986; Hoff et al., 2002; Snijders et al., 2006). Therefore, power calculations for HNMs are not as straightforward as calculations for HLMs since we cannot assume independence among ties.

Second, we focus on the relationship between the standard error and sample size. In a HLM, we see that the number of clusters affects standard error more than cluster size in that only number of clusters can drive the standard error to zero. In fact, we will show that in HNMs, cluster size is actually more influential than number of clusters and that both can drive standard error to zero.

3. Standard Error Calculations for the HLSM

Power calculations hinge on an accurate estimate of the intervention effect standard error. Because the HNM dependence structure differs from classical HLM dependence, standard error calculations for HLMs are not applicable. Instead we develop consistency, asymptotic normality and standard error results from first principles.

Consider an experiment on a collection of networks. In a typical experiment, some networks receive the treatment and the remainder are considered control networks. For example, if each network is a school, we might randomly assign schools to be in the treatment or control arms of our experiment and we assume the schools to be independent in that the outcome of one school is not influenced or dependent on the outcome of another school. We also recognize that school networks are not identical in size, structure or functionality, and our statistical models reflect that. In the HNM framework, we model these networks to be independent of one another but not identically distributed.

Note that consistency of MLEs when data is independently and identically distributed is well documented (Casella & Berger, 2002). Given independent but *not* identically distributed data, Cramér (1946) proved that the maximum likelihood estimate is both consistent and asymptotically normal. We adapt his work along with other adaptations of Chanda (1954) and Bradley & Gart (1962) to show that the posterior mode estimator $\hat{\theta}$ in (1) is both consistent and asymptotically normal. We then use the asymptotic distribution for $\hat{\theta}$ to estimate $SE(\hat{\theta})$ and compute power.

Our development is organized into two main theorems. The first shows that $\hat{\theta}$ in (1) is consistent. The second shows asymptotic normality. For each theorem, we first establish the result conditional on the latent space positions $\mathbb{Z}$ and then prove the result holds unconditional on the latent space positions. Formally, our results are conditional on fixing the coefficients β or any other coefficients of W in (1). However, following arguments similar to Theorems 4.3 and 4.5, the same results hold unconditional on β.

3.1 Consistency of the Posterior Mode $\hat{\theta}$

For all results presented in this section, we use the following notation. For network k with m individuals, the HLSM is given as

$$\begin{aligned} \text{logit}\left[P(Y_{ijk}=1)\right] &= \beta_0 + \beta_1 W_{ijk} + \theta X_k - d(Z_{ik}, Z_{jk}) \\ &= \beta W_{ijk} + \theta X_k - d_{ijk}\ , i,j = 1,\dots,m_k \end{aligned} \tag{3}$$

where θ is the treatment effect and X_k is the treatment group indicator. Denote $\mathbb{Z}$ as the set of latent space positions and let d_{ijk} be the distance between latent space positions Z_{ik} and Z_{jk}. By augmenting W with a column of ones and collecting $\beta = (\beta_0, \beta_1)$, we write βW_{ijk} for $\beta_0 + \beta_1 W_{ijk}$ in the remainder of the paper.

Furthermore, let $L_k(Y_k, \theta)$ be the probability density function for each independent network, and the likelihood for ties from n networks be written as $\prod_{k=1}^{n} L_k(Y_k, \theta)$. Given the prior for θ as $\pi(\theta)$, we denote the posterior of θ, $p = p(\theta|Y_1,\dots,Y_k) \propto \pi(\theta)\prod_{k=1}^{n} L_k(Y_k,\theta)$ with mode $\hat{\theta}$. Note also that the support of Y_k is $\{0,1\}$ since ties are either absent or present.

The log posterior and its first, second, and third derivatives are given as

$$\begin{aligned} \log p &= \log\left(\pi(\theta)\right) + \sum_{k=1}^{n}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k} Y_{ijk}(\beta W_{ijk} + \theta X_{ijk} - d_{ijk}) - \log\left(1+\exp\left(\beta W_{ijk} + \theta X_{ijk} - d_{ijk}\right)\right) \\ \frac{\partial \log p}{\partial\theta} &= \frac{\partial \log \pi(\theta)}{\partial\theta} + \sum_{k=1}^{n}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k} X_k Y_{ijk} - \frac{X_k \exp\left(\beta W_{ijk} + \theta X_k - d_{ijk}\right)}{1+\exp\left(\beta W_{ijk} + \theta X_k - d_{ijk}\right)} \\ \frac{\partial^2 \log p}{\partial\theta^2} &= \frac{\partial^2 \log \pi(\theta)}{\partial\theta^2} - \sum_{k=1}^{n}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k} \frac{X_k^2 \exp\left(\beta W_{ijk} + \theta X_k - d_{ijk}\right)}{\left(1+\exp\left(\beta W_{ijk} + \theta X_k - d_{ijk}\right)\right)^2} \\ \frac{\partial^3 \log p}{\partial\theta^3} &= \frac{\partial^3 \log \pi(\theta)}{\partial\theta^3} + \sum_{k=1}^{n}\sum_{\substack{i,j=1\\ i\neq j}}^{m_k} \frac{X_k^3 \exp\left(\beta W_{ijk} + \theta X_k - d_{ijk}\right)\left(\exp\left(\beta W_{ijk} + \theta X_k - d_{ijk}\right)-1\right)}{\left(1+\exp\left(\beta W_{ijk} + \theta X_k - d_{ijk}\right)\right)^3}\ . \end{aligned} \tag{4}$$

We begin by verifying some results needed for Theorem 1. Let $\xrightarrow{p}$ denote convergence in probability.

Proposition 1. Let θ_0 be the true treatment effect. Suppose $\frac{\partial \log \pi}{\partial \theta}$ and its derivatives are bounded and suppose $m_k \leq m^\star < \infty$, $\forall k = 1, \ldots, n$. Then,

1. $\frac{1}{n}\frac{\partial \log p}{\partial \theta}\big|_{\theta=\theta_0} \xrightarrow{p} 0$ as $n \longrightarrow \infty$
2. $\frac{1}{n}\frac{\partial^2 \log p}{\partial \theta^2}\big|_{\theta=\theta_0} + \frac{1}{n}\sum_{k=1}^{n} C_k^2 \xrightarrow{p} 0$ as $n \longrightarrow \infty$ where

$$C_k^2 = E\left[\sum_{\substack{i,j=1\\ i\neq j}}^{m_k} \frac{X_k^2 \exp\left(\beta W_{ijk} + \theta_0 X_k - d_{ijk}\right)}{\left(1+\exp\left(\beta W_{ijk} + \theta_0 X_k - d_{ijk}\right)\right)^2}\right] .$$

Proof. The derivatives $\frac{\partial \log p}{\partial \theta}$ and $\frac{\partial^2 \log p}{\partial \theta^2}$ are sums of at most $nm^\star(m^\star - 1)$ finite terms plus the log prior derivative term which by assumption converges to 0 as n grows. By applying a version of the Weak Law of Large Numbers, (Kintchine's Theorem Cramér, 1946, p. 253), the Proposition is proved. □

Note that $C_k^2 > 0$ when $X_k = 1$, i.e. when the network is in the treated condition. In practice, we expect the number of treated networks to grow as n grows and in fact to be on the order of $O(n)$ and our results rely on this assumption.

Lemma 1. ***(Consistency of the Posterior Mode, Conditional on the Latent Space Positions)***

Let θ be the treatment effect parameter and let $\hat{\theta}$ be the value that maximizes the posterior distribution and let θ_0 be the true effect. Let $X_k > 0$ at least n' times where n' is on the order of $O(n)$. Then for an arbitrarily $\epsilon > 0$,

$$P\left(|\hat{\theta} - \theta_0| > \epsilon|\mathbb{Z}\right) \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Proof. To show consistency, we first represent the log posterior of θ using a second order Taylor approximation, and we then show convergence for each term. Using these results, we then prove that a solution to the normal equation $\dfrac{\partial \log p}{\partial \theta} = 0$ exists, is a maximum, and is arbitrarily close to θ_0.

Let $\delta > 0$ be small. Then in a neighborhood of θ_0, we can approximate the derivative of the log posterior as a second-order Taylor polynomial,

$$\frac{1}{n}\frac{\partial \log p}{\partial \theta} = \frac{1}{n}\frac{\partial \log p}{\partial \theta}\Big|_{\theta=\theta_0} + (\theta-\theta_0)\frac{1}{n}\frac{\partial^2 \log p}{\partial \theta^2}\Big|_{\theta=\theta_0} + \frac{1}{2}(\theta-\theta_0)^2\frac{1}{n}\frac{\partial^3 \log p}{\partial \theta^3} \tag{5}$$

$$= \frac{1}{n}\frac{\partial \log p}{\partial \theta}\Big|_{\theta=\theta_0} + (\theta-\theta_0)\frac{1}{n}\frac{\partial^2 \log p}{\partial \theta^2}\Big|_{\theta=\theta_0} + \frac{\lambda}{2}(\theta-\theta_0)^2\frac{1}{n}M \tag{6}$$

Note that $\dfrac{\partial^3 \log p}{\partial \theta^3}$ is bounded so there exists some constant $M < \infty$ such that $\left|\dfrac{\partial^3 \log p}{\partial \theta^3}\right| < M$ and $\dfrac{\partial^3 \log p}{\partial \theta^3} = \lambda M$ for some $|\lambda| < 1$. Then (5) and (6) are equivalent.

Denoting $B_0 = \dfrac{1}{n}\dfrac{\partial \log p}{\partial \theta}\Big|_{\theta=\theta_0}$ and $B_1 = \dfrac{1}{n}\dfrac{\partial^2 \log p}{\partial \theta^2}\Big|_{\theta=\theta_0}$, we can write (6) in terms of B_k,

$$\frac{1}{n}\frac{\partial \log p}{\partial \theta} = B_0 + (\theta-\theta_0)B_1 + \frac{\lambda}{2}(\theta-\theta_0)^2 M \tag{7}$$

Let $\delta > 0$ and $\epsilon > 0$ be fixed, and consider evaluating $\dfrac{\partial \log p}{\partial \theta}$ in a neighborhood of θ_0, $(\theta_0 - \delta, \theta_0 + \delta)$. We obtain the two equations

$$\frac{1}{n}\frac{\partial \log p}{\partial \theta}\Bigg|_{\theta-\delta} = B_0 - \delta B_1 + \frac{\lambda}{2}\delta^2 M = 0 \tag{8}$$

$$\frac{1}{n}\frac{\partial \log p}{\partial \theta}\Bigg|_{\theta+\delta} = B_0 + \delta B_1 + \frac{\lambda}{2}\delta^2 M = 0 \ , \tag{9}$$

for $\theta_0 - \delta$ and $\theta_0 + \delta$ respectively. We now show that (8) is positive and (9) is negative.

By Proposition 1, the following is true for sufficiently large n:

- $|B_0| < \delta^2$
- $B_1 + \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} C_k^2 < \mu$ for some $\mu > 0$.

Since $C_k^2 > 0$ for $O(n)$ networks, there exists some n_0 such that $B_1 > \frac{1}{2n}\sum_{k-1}^{n} C_k^2$ for all $n > n_0$. Then let n be sufficiently large, so that

$$\begin{aligned} B_0 + \frac{\lambda}{2}\delta^2 M &< \delta^2 + \frac{\lambda}{2}\delta^2 M \\ &< \delta^2(1 + \frac{M}{2}) \\ &< \frac{\delta}{2n}\sum_{k=1}^{n} C_k^2 \\ &< \delta B_1 \ , \end{aligned}$$

since for large n, $\delta < \frac{\frac{1}{n}\sum_{k=1}^{n} C_k^2}{2+M}$. Thus, the B_1 term controls the sign of (7). Because $B_1 < 0$, it follows that $\frac{\partial \log p_k}{d\theta}$ is positive when evaluated at $(\theta_0 - \delta)$ and negative at $(\theta_0 + \delta)$. Therefore, there exists a solution to the equation $\frac{\partial \log p_k}{d\theta} = 0$ that is arbitrarily close to θ_0 and is a maximum. Note that this is the only maximum and is therefore the absolute maximum of the log posterior which we defined to be $\hat{\theta}$.

Thus, the posterior mode $\hat{\theta}$ is arbitrarily close to the true value θ_0 for adequately large n. Therefore, $P\left(|\hat{\theta} - \theta_0| > \epsilon|\mathbb{Z}\right) \longrightarrow 0$. □

Notice that Theorem 1 assumes the latent space positions are known. We now show that the posterior mode is a consistent estimator, unconditional on the latent space positions. We present this result as a second theorem.

Theorem 1. ***(Consistency of the Posterior Mode, Unconditional on the Latent Space Positions)***

Let θ be the treatment effect parameter and $\hat{\theta}$ be posterior mode where θ_0 is the true effect. Let m_k be the number of nodes in network k such that $\lim_{n\longrightarrow\infty} max(m_1, \ldots, m_n) < \infty$. Then, unconditional on the latent space positions, $\hat{\theta}$ is consistent: $P[|\hat{\theta} - \theta_0| > \epsilon] \longrightarrow 0$ as $n \longrightarrow \infty$.

Proof. From Theorem 1, we have $P\left[\left|\hat{\theta} - \theta_0\right| > \epsilon\middle|\mathbb{Z}\right] \longrightarrow 0$ with measure 1 and $P\left[\left|\hat{\theta} - \theta_0\right| > \epsilon\middle|\mathbb{Z}\right]$ is a probability bounded by 1. Then by the Dominated Convergence Theorem for Expectation (Ash & Doleans-Dade, 2000),

$$E_{\mathbb{Z}}\left[P\left[|\hat{\theta}-\theta_0|>\epsilon|\mathbb{Z}\right]\right] \longrightarrow E_{\mathbb{Z}}\left[0\right]$$
$$P\left[|\hat{\theta}-\theta_0|>\epsilon\right] \longrightarrow 0$$

Thus, we have unconditional consistency of the posterior mode: $\lim_{n\to\infty} P\left(|\hat{\theta}-\theta_0|>\epsilon\right)=0.$ □

3.2 Asymptotic Normality

We now show that the posterior mode $\hat{\theta}$ is asymptotically normal. Again, we present two arguments, normality conditional on the latent space positions and normality unconditional on the positions.

Lemma 2. ***Asymptotic Normality of the Posterior Mode, Conditional on the Latent Space Positions***

Let θ be the treatment effect in a HLSM with posterior mode $\hat{\theta}$ and true effect θ_0. Let π be the prior distribution for θ and let $\frac{\partial \pi}{\partial \theta}$ be bounded. Then conditional on the latent space positions, $\hat{\theta}$ is asymptotically normal, i.e. $\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \xrightarrow{D} N(0,1)$, where C_k^2 is the Fisher Information for network k and is greater than 0 for $O(n)$ networks.

Proof. Using notation introduced in Lemma 1 and (7), the normal equation can be rewritten as:

$$\begin{aligned}
&B_0 + (\hat{\theta}-\theta_0)B_1 + \frac{1}{2}B_2\lambda(\hat{\theta}-\theta_0)^2 = 0 \\
&\implies (\hat{\theta}-\theta_0) = \frac{\frac{1}{n}\left(\frac{\partial \log p}{\partial \theta}|_{\theta=\theta_0}\right)}{-B_1 - \frac{1}{2}\lambda B_2 \lambda(\hat{\theta}-\theta_0)} \\
&\implies \sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) = \frac{\frac{1}{\sqrt{\sum_{k=1}^{n} C_k^2}}\left(\frac{\partial \log p}{\partial \theta}|_{\theta=\theta_0}\right)}{\frac{n}{\sum_{k=1}^{n} C_k^2}(-B_1 - \frac{1}{2}\lambda B_2(\hat{\theta}-\theta_0))}
\end{aligned} \tag{10}$$

We prove that $\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \longrightarrow N(0,1)$ by showing the right hand side of (10) converges to a normal (0,1) distribution. We first show that $\frac{1}{\sqrt{\sum_{k=1}^{n} C_k^2}}\sum_{i=1}^{n}\left(\frac{\partial \log p}{\partial \theta}|_{\theta=\theta_0}\right) \longrightarrow N(0,1)$. We then prove that $\frac{n}{\sum_{k=1}^{n} C_k^2}(-B_1 - \frac{1}{2}\lambda B_2(\hat{\theta} - \theta_0))$ converges to 1 and quick application of Slutsky concludes our proof.

Let us write $\frac{\partial \log p}{\partial \theta}\Big|_{\theta=\theta_0} = \frac{\partial \log \pi}{\partial \theta}\Big|_{\theta=\theta_0} + \sum_{k=1}^{n} \ell_k$, where $\ell_k = \sum_{i,j}^{m} \left[X_k Y_{ijk} - \frac{X_k \exp\left(\beta W_{ijk} + \theta_0 X_k - d_{ijk}\right)}{1 + \exp\left(\beta W_{ijk} + \theta_0 X_k - d_{ijk}\right)} \right]$. Then clearly, $\lim_{n\to\infty} \frac{1}{\sqrt{\sum_{k=1}^{n} C_k^2}} \frac{\partial \log \pi}{\partial \theta}\Big|_{\theta=\theta_0} = 0$ since $C_k^2 > 0$ for $O(n)$ networks and $\frac{\partial \log \pi}{\partial \theta}$ is bounded.

We now use the Lyapunov Central Limit Theorem (Ash & Doleans-Dade, 2000) to show that $\frac{1}{\sqrt{\sum_{k=1}^{n} C_k^2}} \sum_{k=1}^{n} \ell_k \longrightarrow N(0,1)$. Since expectation $E[\ell_k] = 0$ and variance $V[\ell_k] = C_k^2$ are bounded, we need only show that for some $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{\left(\sqrt{\sum_{k=1}^{n} C_k^2}\right)^{2+\delta}} \sum_{k=1}^{n} E\left[|\ell_k|^{2+\delta}\right] = 0 \ . \tag{11}$$

First, both C_k and ℓ_k are bounded. Second, notice that when $X_k = 0$ both C_k^2 and ℓ_k are 0. Thus, when $X_k = 1$, there exists some $L < \infty$ such that $\ell_k < m^\star L$ and there exists some $C > 0$ such that $C < C_k^2$ for all $k = 1 \ldots n$. Then (11) $< \frac{n(Lm^\star)^{2+\delta}}{(nC)^{1+\delta/2}}$ which converges to 0 as $n \longrightarrow \infty$.

We now prove that $\frac{n}{\sum_{k=1}^{n} C_k^2}(-B_1 - \frac{1}{2}\lambda B_2(\hat{\theta} - \theta_0))$ converges to 1. By Theorem 1 and the Continuous Mapping Theorem (Ash & Doleans-Dade, 2000), $C(\hat{\theta} - \theta_0) \xrightarrow{p} 0$ for any constant C. We also have the result that $B_1 - (-\frac{1}{n}\sum_{i=1}^{n} C_k^2) \xrightarrow{p} 0$, so we claim that $\frac{-B_1}{\frac{1}{n}\sum_{k=1}^{n} C_k^2} - 1 \xrightarrow{d} 0$.

By assumption, C_k^2 is strictly positive for $O(n)$ terms, so that the limit is strictly greater than zero, $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} C_k^2 > 0$. And by definition, $C_k^2 = \sum_{i,j}^{m_k} \frac{X_k^2 \exp\left(\theta X_k - d_{ijk}\right)}{(1 + \exp\left(\theta X_k - d_{ijk}\right))^2}$ which is a sum of at most $m^\star(m^\star - 1)$ terms each bounded by 0.25. Then, $\frac{1}{n}\sum_{k=1}^{n} C_k^2 < \frac{m^\star(m^\star - 1)}{4} < \infty$. Thus, by Slutsky's Theorem (Ash & Doleans-Dade, 2000), $\frac{-B_1}{\frac{1}{n}\sum_{k=1}^{n} C_k^2} - 1 \xrightarrow{d} 0$ and the entire denominator in (10) converges to 1.

Thus, $\sum_{k=1}^{n} \left(\frac{\partial \log p_k}{\partial \theta}|_{\theta=\theta_0}\right) \xrightarrow{D} N(0, \sum_{k=1}^{n} C_k^2)$.

And by Slutsky, $\frac{\frac{1}{\sum_{k=1}^{n} C_k^2}\sum_{i=1}^{n}\left(\frac{\partial \log p_k}{\partial \theta}|_{\theta=\theta_0}\right)}{\frac{n}{\sum_{k=1}^{n} C_k^2}(-B_1 - \frac{1}{2}\lambda B_2 \lambda(\hat{\theta} - \theta_0))} \xrightarrow{D} N(0,1)$.

Therefore, $\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \xrightarrow{D} N(0,1)$, conditional on the latent space positions $\mathbb{Z}$. □

Theorem 2. ***Asymptotic Normality of the Posterior Mode, Unconditional on the Latent Space Positions***

Let θ be the treatment effect in a HLSM with posterior mode $\hat{\theta}$ and true effect θ_0. Then **unconditional on the latent space positions**, $\hat{\theta}$ is asymptotically normal, i.e. $\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \xrightarrow{D} N(0,1)$, where C_k^2 is the Fisher Information of θ for network k.

Proof. By Theorem 2, $\lim_{n\to\infty} P_{Y|\mathbb{Z}}\left[\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \leq t|\mathbb{Z}\right] = \Phi(t),\ \forall t.$

Let $Y_n = P_{Y|\mathbb{Z}}\left[\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \leq t|\mathbb{Z}\right]$ and $Y = \Phi(t)$ for any t. Again Y_n is a probability and is bounded by 1. By the Dominated Convergence Theorem for Expectation(Ash & Doleans-Dade, 2000) and using the fact that $\mathbb{Z}$ is independent of t,

$$E\left[P_{Y|\mathbb{Z}}\left[\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \leq t|\mathbb{Z}\right]\right] \longrightarrow E[\Phi(t)],\ \forall t$$

$$P_Y\left[\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \leq t\right] \longrightarrow \Phi(t),\ \forall t.$$

Thus, we have asymptotic normality unconditional on the latent space positions,

$P_Y\left[\sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \leq t\right] \longrightarrow \Phi(t),\ \forall t \iff \sqrt{\sum_{k=1}^{n} C_k^2}(\hat{\theta}-\theta_0) \xrightarrow{D} N(0,1).$ □

3.3 Power Intervals

Given the asymptotic distribution of $\hat{\theta}$, $\hat{\theta} \xrightarrow{D} N\Big(\theta_0, \frac{1}{\sum_{k=1}^{n} C_k^2}\Big)$, we can use $\sqrt{\frac{1}{\sum_{k=1}^{n} C_k^2}}$ as an estimate for $SE(\hat{\theta})$ and calculate power.

Recall from $\sum_{k=1}^{n} C_k^2$ is a function of the latent space positions which are not known *a priori*. Instead, we find reasonable bounds for $\sum_{k=1}^{n} C_k^2$, taking advantage of the log-odds ratio structure of the HLSM,

$$\sum_{k=1}^{n} C_k^2 = \sum_{k=1}^{n} \sum_{i,j}^{m_k} \Big[\frac{X_k^2 \exp(\theta X_k - d_{ijk})}{(1+\exp(\theta X_k - d_{ijk}))^2}\Big] \ .$$

Without loss of generality, e.g. by subsuming βW_{ijk} into $-d_{ijk}$, suppose $\beta W_{ijk} = 0, \ \forall i,j,k$ and there is constant network size m. We can easily construct an upper bound for $\sum_{k=1}^{n} C_k^2$ since $\frac{X_k^2 \exp(\theta X_k - d_{ijk})}{(1+\exp(\theta X_k - d_{ijk}))^2}$ has a maximum possible value of $\frac{1}{4}$ for any value of $X_k \in \{0,1\}$, $\theta \in \mathbb{R}$, and $d_{ijk} \in \mathbb{R}$. Therefore,

$$\sum_{k=1}^{n} C_k^2 \leq \frac{mn(m-1)}{4},$$

where m is the number of nodes or members per network and n is the number of networks.

For a lower bound, if we assume $|\theta - d_{ijk}| \leq a$ then $\frac{\exp(\theta - d_{ijk})}{(1+\exp(\theta - d_{ijk}))^2} \geq \frac{\exp(a)}{(1+\exp(a))^2}$, for all values $|\theta - d_{ijk}| < a$. And in fact, $\frac{\exp(\theta - d_{ijk})}{(1+\exp(\theta - d_{ijk}))^2}$ asymptotically approaches 0 as $|\theta - d_{ijk}| \longrightarrow \infty$. Figure 2 illustrates this relationship.

Let f be a lower bound on the proportion of treated networks ($X_k = 1$), so that

$$\sum_{k=1}^{n} C_k^2 = \sum_{k=1}^{n} \sum_{i,j}^{m} \Big[\frac{X_k^2 \exp(\theta X_k - d_{ijk})}{(1+\exp(\theta X_k - d_{ijk}))^2}\Big] \geq \frac{fnm(m-1)\exp(a)}{(1+\exp(a))^2} \ .$$

Therefore, we have the following bounds:

$$\frac{fnm(m-1)e^a}{(1+e^a)^2} \leq \sum_{k=1}^{n} C_k^2 \leq \frac{nm(m-1)}{4} \ ,$$

and since $Var(\theta) = \frac{1}{\sum_{k=1}^{n} C_k^2}$,

$$\frac{4}{nm(m-1)} \leq \mathrm{Var}(\hat{\theta}) \leq \frac{(1+e^a)^2}{fnm(m-1)e^a} \ . \tag{12}$$

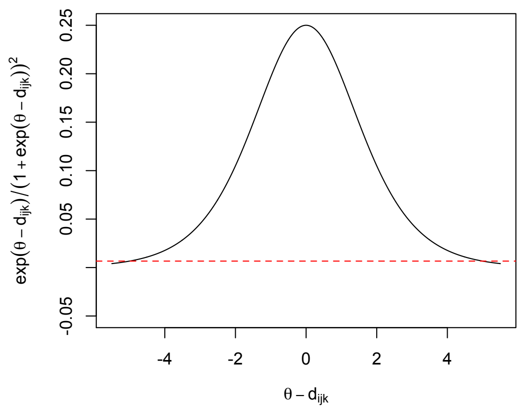


FIGURE 2.
Relationship between $\dfrac{\exp(\theta - d_{ijk})}{(1+\exp(\theta - d_{ijk}))^2}$ and $\theta - d_{ijk}$.

From (12) we see that the number of networks n and the number of ties $m(m-1)$ drive standard error equally. Thus the network size m has greater influence over standard error than network number n which is not the case for a standard HLM, in which n drives error.

4. Power Calculations

Given the bounds for $\mathrm{Var}(\hat{\theta})$, we can easily construct upper and lower bounds for power, given appropriate choices for f and a. In this section, we discuss methods for selecting a and f as well as calculate bounds for experiments involving networks of various size and intervention effect.

Recall that f is a lower bound on the proportion of treated networks. Since an experiment hinges on having treated subjects, $f > 0$ and typically $f = \dfrac{1}{2}$. Choosing a is more complicated since $|\theta - d_{ijk}| \leq a$ corresponds to the probability of a tie being in the interval $\left[\dfrac{\exp(-a)}{1+\exp(-a)}, \dfrac{\exp(a)}{1+\exp(a)}\right]$, so by choosing a value for a, we are choosing the range of tie probabilities. Taking advantage of the logit structure of the HLSM, a large value of a increases the range of probabilities only slightly more than a moderately large value of a but large differences in a result in large differences in power estimates as we will later demonstrate. To provide a context, we compute the ranges of tie probabilities when $a = 4$, $a = 5$, and $a = 6$ (Table 1).

All three values of a seem plausible based on the wide intervals of tie probabilities, and we

TABLE 1.

The interval of tie probabilities specified by the choice of a where $|\theta - d_{ijk}| \leq a$ and the probability of a tie is $\frac{\exp(\theta - d_{ijk})}{1 + \exp(\theta - d_{ijk})}$, see Equation 3.

Probability Bounds	a
$[0.018, 0.982]$	4
$[0.007, 0.993]$	5
$[0.002, 0.997]$	6

will use these three values for comparison as we construct our power intervals, which we calculate using the following formula.

$$\text{Power}_{UB} = P_{H_1}(\hat{\theta} > \theta_{NULL} + Z_{\frac{\alpha}{2}} SE(\hat{\theta})_{LB})$$
$$\text{Power}_{LB} = P_{H_1}(\hat{\theta} > \theta_{NULL} + Z_{\frac{\alpha}{2}} SE(\hat{\theta})_{UB}) \tag{13}$$

Assuming $\theta_{NULL} = 0$ (i.e. no treatment effect), Table 2 provides examples of lower bounds for power for a various combination of numbers of network sizes (members), number of networks in the experiment, two different alternative hypotheses, and our three values of a. Table 2 only includes lower bounds since the upper bound is ≈ 1 for all cells. The upper bound is generally close to 1 because the lower bound for standard error $\frac{4}{nm(m-1)}$ is quite small even for moderate sizes of m and n. However, we consider the lower bound for power to be more important since power calculations are generally used to determine the minimum sample size needed to achieve specific power.

As expected, decreasing effect size decreases power and increasing the number of networks increases power. What is surprising however, is the effect of network size on power. Increasing the network size from 10 to 15 individuals greatly increases power regardless of treatment effect, value for a or number of networks. In fact the number of networks n and the number of possible ties $m(m-1)$ contribute equally to the power calculations since they equally contribute in asymptotically driving $SE(\hat{\theta})$ to zero, see Equation 12. This relationship is not true in a typical HLM power calculation, see Equation 2, since cluster size m does not drive the standard error to zero. Our lower bound for power also depends heavily on the constant a, so it is important to choose a carefully.

Table 2.

Examples of Power Bounds with various combinations of number of networks (10 or 20) and members per network (10 or 15), treatment effect (TE of 0.25 or 0.5) and constant (a= 4, 5, or 6) to illustrate the effects that each have on power. Upper bounds are not shown and are approximately 1.

TE	(m, n)	Power Lower Bound		
		$a = 6$	$a = 5$	$a = 4$
$\theta = 0.5$	(10,10)	0.07	0.13	0.28
	(10,20)	0.10	0.22	0.50
	(15,10)	0.12	0.25	0.56
	(15,20)	0.19	0.44	0.85
$\theta = 0.75$	(10,10)	0.11	0.24	0.55
	(10,20)	0.19	0.43	0.84
	(15,10)	0.21	0.49	0.89
	(15,20)	0.38	0.79	0.99

5. Simulations

We conduct a series of simulations, both to assess the analytical power bound results derived in Section 3.3, as well as to understand how these bounds can be used in practice. We first examine the distribution of the treatment effect estimator, since our results assume asymptotical normality of the estimator. Then, we compare empirical power estimates with the analytical lower and upper bounds.

Regarding practical use, our primary goal is to reveal plausible values for the constant a in Equation 12 and how network size affects the choice of a. It is not unreasonable to assume that the range of plausible probabilities increases as the number of possible ties increase. We use empirical results to shape our understanding of this relationship. We also use empirical power estimates to further explore the relationship between network size and number of networks.

The simulation study involved a combination of 2 treatment effects, 3 network sizes and 6 different total numbers of networks. We specifically chose treatment effects, network sizes and numbers of networks to cover a wide range of power estimates while focusing on network sizes that could be found in schools or classrooms. We assumed equal numbers of treated and untreated networks. The full design is given in Table 3.

Table 3.

Simulation Summary: For each cell marked with an x, we simulated 100 data sets for that combination of m and n.

$\theta = 0.5$		Number of Networks (n)					
		6	10	20	30	40	50
Network	10	x	x	x	x	x	x
Size	15	x	x	x	x	x	x
(m)	20	x	x	x	x	x	x

$\theta = 0.75$		Number of Networks (n)					
		6	10	20	30	40	50
Network	10	x	x	x	x	x	x
Size	15	x	x	x	x	x	x
(m)	20	x	x	x	x	x	x

For each cell in Table 3, we simulated 100 datasets from the generative model

$$\text{logit}P[Y_{ijk} = 1] = \beta_0 + \theta T_k + |Z_{ik} - Z_{jk}|$$

$$Z_{ik} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}\right),\ i = 1, \ldots, m\ , k = 1, \ldots, n\ ,$$

where $\beta_0 = 1.5$ is a fixed intercept.

For each dataset, we fit a similar HLSM to the simulated data using uninformative priors, $\theta \sim N(0, 100)$ and $\beta_0 \sim N(0, 100)$ using an MCMC algorithm coded in R (Sweet et al., 2013). For each model fit, we discard the first 1000 samples, and retain every 25^{th} draw from our remaining 9000 samples for a θ posterior sample of 360.

Because our power calculations are based on the asymptotic normal distribution of the posterior mode $\hat{\theta}$, we compare the distribution of the 100 posterior modes to that of a normal distribution. Visual inspection of Q-Q plots suggests that the posterior modes are approximately normal (Figure 3).

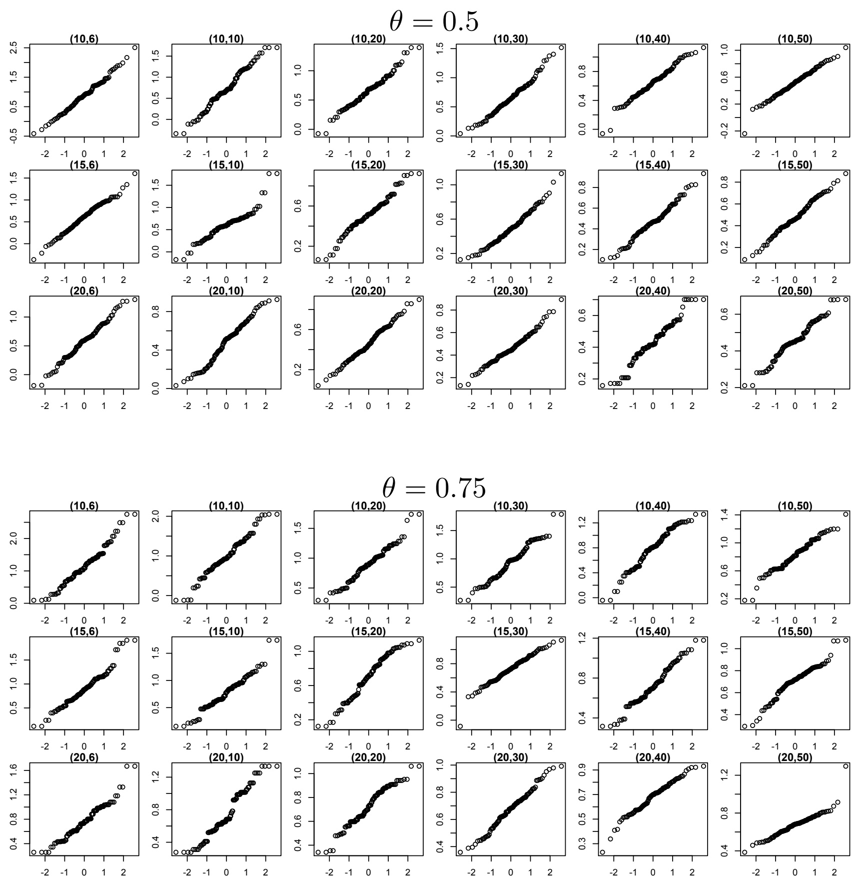


FIGURE 3.

Q-Q plots for the 100 replications of posterior mode in the simulation study. The plots are labeled as (network size, number of networks). The posterior mode appears to be approximately normal.

5.1 Empirical Power Estimates

To evaluate our power bound formulas, we compare our lower bound estimates with empirical power estimates. For each combination of (θ, m, n) shown in Table 3, we estimate power by taking the proportion of trials with detectable treatment effects, i.e. the 95% equal–tailed credible interval for θ did not include 0.

Table 4 compares the relationship between the calculated lower bounds calculated using three different values of a and simulated power estimates. Regardless of treatment effect, increasing the number or size of networks generally increases power, and the larger treatment effect generates larger power estimates, as predicted by our formulas in (12) and (13) respectively. We also include an estimate for level, i.e. when the true treatment effect is 0.

Empirical estimates also substantiate the claim that the size of each network m contributes more to power than the number of networks n through error as shown in (12). For example, when the true treatment effect is 0.5, the power estimate for an experiment on 10 networks of size 10 each with 10 individuals is 0.36. Increasing the number of networks to 20 yields estimated power of 0.48 whereas increasing the network size to 20 yields power of 0.60. This pattern of network size influencing power more than network number is also true when the treatment effect is 0.75.

This relationship between the number of possible ties, the number of networks and power is apparent in Figure 4 which illustrates the analytical lower bounds calculated and the power recovered by the simulation studies. The x-axis is the total number of ties across all networks, $nm(m-1)$, and the simulated power estimates follow a typical power curve, again validating that the number of possible ties and number of networks equally drive power.

Figure 4 also illustrates how a in (12) influences the analytical power estimates and how these compare with empirical power estimates. It appears that a value between 4 and 5 is most appropriate. We also note that Figure 4 suggests that the value of a may differ by network size, which is not surprising. Larger networks have more ties and more opportunities to generate tie probabilities that are either very small or very large.

To determine whether a varies with network size in practice, we use the simulated data to empirically estimate a. For each cell of our simulation study, we have 100 replications of n

Table 4.

Power lower bounds compared to the empirical estimate for power based on 100 simulations for treatment effect $\theta = 0.5$ and $\theta = 0.75$ where m is the network size and n is the number of networks. The power estimates for $\theta = 0$ are also shown for comparison. Increasing the network size makes improves power more than increasing number of networks. Upper bound for power is approximately 1 for most cells and is not shown.

Treatment	$\theta = 0$	$\theta = 0.5$				$\theta = 0.75$			
	Level	Lower Bound			Power	Lower Bound			Power
(m,n)	Est.	a=6	a=5	a=4	Est.	a=6	a=5	a=4	Est.
(10, 6)	0.09	0.06	0.09	0.28	0.35	0.08	0.16	0.36	0.54
(10, 10)	0.05	0.07	0.13	0.28	0.36	0.11	0.24	0.55	0.64
(10, 20)	0.08	0.10	0.22	0.50	0.48	0.19	0.43	0.84	0.81
(10, 30)	0.07	0.14	0.31	0.67	0.69	0.26	0.60	0.95	0.85
(10, 40)	0.06	0.17	0.39	0.79	0.78	0.34	0.72	0.99	0.83
(10, 50)	0.03	0.21	0.47	0.88	0.74	0.41	0.82	≈ 1	0.98
(15, 6)	0.09	0.08	0.16	0.37	0.34	0.14	0.32	0.69	0.60
(15, 10)	0.07	0.12	0.25	0.56	0.52	0.21	0.49	0.89	0.71
(15, 20)	0.08	0.19	0.45	0.85	0.79	0.38	0.79	0.99	0.81
(15, 30)	0.13	0.27	0.61	0.96	0.80	0.53	0.92	≈ 1	0.98
(15, 40)	0.08	0.35	0.74	0.99	0.82	0.66	0.97	≈ 1	1.00
(15, 50)	0.09	0.42	0.83	≈ 1	0.91	0.76	0.99	≈ 1	1.00
(20, 6)	0.08	0.12	0.27	0.60	0.54	0.23	0.53	0.91	0.80
(20, 10)	0.03	0.18	0.41	0.82	0.60	0.35	0.75	0.99	0.83
(20, 20)	0.10	0.32	0.70	0.98	0.81	0.62	0.96	≈ 1	1.00
(20, 30)	0.09	0.45	0.86	≈ 1	0.95	0.79	≈ 1	≈ 1	1.00
(20, 40)	0.04	0.57	0.94	≈ 1	0.99	0.89	≈ 1	≈ 1	1.00
(20, 50)	0.07	0.6	0.98	≈ 1	1.00	0.95	≈ 1	≈ 1	1.00

networks, each with $M = m(m-1)$ ties. Within each replication and for each network, we examined the absolute probabilities of each tie, $|\beta_0 + \theta X_k - d_{ijk}|$, and use the 99^{th} quantile as the bound for that network. We then use the mean bound across all n networks and 100 replications for an empirical estimate of a. Thus, $\hat{a} = \frac{1}{100}\sum_{s=1}^{100}\frac{1}{n}\sum_{k=1}^{n} a^{\star}_{ks}$ where $a^{\star}_{ks} = |\beta_0 + \theta X_k - d_{ijk}|_{0.99}$ for all i, j in network k and simulation s.

Table 5 illustrates the relationship between network size and a. In fact, even for small increases in network size from 10 to 15 to 20, we see small but distinct increases in a.

TABLE 5.
Empirical estimate for the size of a, the bound on the tie probability, calculated from simulated data.

n	Network of Size 10	Network of Size 15	Network of Size 20
6	4.43	4.73	4.96
10	4.38	4.71	4.96
20	4.38	4.74	4.97
30	4.37	4.77	4.94
40	4.34	4.74	4.94
50	4.35	4.75	4.95

Our simulation study supports our analytical result that power is driven by both network size and the number of networks and that increasing the size of the networks does more to improve power than increasing the number of networks. These simulations also suggest possible values for a. We find empirically that $a = 5$ is an appropriate upper bound for $|\theta - d_{ijk}|$ although our simulations also suggest an additional caveat that a is a function of network size. In fact, using the empirical estimates generated by Table 5 of $a = (4.5, 4.8, 5)$ for network sizes $m = (10, 15, 20)$, one may observe in Figure 4 that the analytical lower bound appears to be both accurate and fairly precise.

6. Discussion

Interventions whose aim is to shape the social structure of a school should use social network data to detect effects and measure relational outcomes. Changes to the network and organizational structure are the first detectable result and can provide insight into whether and how the intervention was successful. Some statistical network models have been proposed for analyzing experimental network data, but there is a void in research on experimental design involving networks.

In this paper, we have introduced the first formulas relating sample size and power for experiments on independent, isolated networks. To derive these formulas, we developed novel results regarding the posterior mode estimator for a network-level treatment effect in a simple hierarchical latent space model; we proved that this estimator is both consistent and follows an asymptotic normal distribution centered at the true treatment effect.

We then used these results to derive bounds for power which are based on the number of networks, the size of each network, and the treatment effect. Most surprising is that the number of possible ties and the number of networks contribute equally to power. We also ran a small simulation study to assess these intervals and the empirical estimates substantiate our analytical results as well as suggested additional relationships. Of the bounds, the lower bound is most practical since it is not only the more precise bound but it is also the one useful for calculating minimum sample size needed to attain a predetermined power level. The simulations also suggest that network size influences the constant a, larger networks require larger values of a and empirically estimating a found a similar relationship. Finally, the empirical estimates do substantiate our analytical results. We find increasing either network size, number of networks, or treatment effect to improve power, and we found a similar relationship between sample size and power in that increasing the network size increases power more than a similar increase in number of networks.

This paper provides a novel yet solid foundation and direction for sample size calculations for network-level experiments. As the first of its kind, there are several areas for future exploration. We have presented results for a single multi-level statistical network model, but there are other models that can accommodate interventions (Sweet, 2012) and we aim to develop similar results for other models. Furthermore, we will also consider ways of narrowing these power bounds; including baseline covariates generally improves power in other models and we believe covariates may be useful in network models as well. We are also considering alternative approaches for sample calculations as well. Finally, empirical results were very useful in refining this work and we believe simulations will continue to serve in both exploratory and confirmatory capacities to forward this work.

References

Ash, R. & Doleans-Dade, C. (2000). *Probability and measure theory*. San Diego, CA: Academic Press.

Bradley, R. & Gart, J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika*, *49*(1/2), 205–214.

Casella, G. & Berger, R. (2002). *Statistical inference*. Pacific Grove, CA: Duxbury Press.

Chanda, K. (1954). A note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, *41*(1/2), 56–61.

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.

Espelage, D. & Low, S. (2009). Multi-site evaluation of second step: Student success through prevention (second step - sstp) in preventing bullying and sexual violence. CDC Grant.

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

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* (pp. 517–563). Santa Monica, CA: RAND Corporation.

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.

Matsumura, L., Garnier, H., & Resnick, L. (2010). Implementing literacy coaching: The role of school social resources. *Educational Evaluation and Policy Analysis*, *32*(2), 249–272.

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

Schervish, M. J. (1997). *Theory of statistics*. New York, NY: Springer.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specification for exponential random graph models. *Sociological Methodology*, *36*(1), 99–153.

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

Sweet, T. (2012). *Staistical Network Models for Replications and Experimental Interventions.* (Unpublished doctoral dissertation). Carnegie Mellon University, Pittsburgh, PA.

Sweet, T. M., Thomas, A. C., & Junker, B. W. (2012). *Modeling Intervention Effects on Subgroup Structure Across Social Networks.* Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

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

Sweet, T. M., Thomas, A. C., & Junker, B. W. (2014). Hierarchical mixed membership stochastic blockmodels for multiple networks and experimental interventions. In E. Airoldi, D. Blei, E. Erosheva, & S. Fienberg (Eds.), *Handbook on Mixed Membership Models and Their Applications*. Boca Raton, FL: Chapman & Hall/CRC.

FIGURE 4.

Power versus Sample Size calculated as the product of the number of networks n, and the number of possible ties $m(m-1)$. Power lower bounds for $a = (3.8, 4, 4.2, 4.4, 4.6, 4.8, 5)$ plotted as curves and simulated power estimates as points. Simulations appear to follow power curve, suggesting network size drives power more than number of networks, i.e. number of possible ties and number of networks drive power equally. Plots also suggest the value of a may be a function of network size.