Supplementary Information for

Consistently Estimating Network Statistics Using Aggregated Relational Data

Emily Breza, Arun G. Chandrasekhar, Shane Lubold, Tyler McCormick, and Mengjie Pan

Tyler McCormick
E-mail:tylermc@uw.edu

This PDF file includes:

We now outline the main parts of the supplementary materials. In Section S1, we provide proofs of Theorems 1 and 2 in the main paper, which deal with consistency in the beta-model and the stochastic block model (SBM), respectively. We then move to proving Theorem 3, which deals with consistency in the latent space model. First, Section S2 defines the estimates of the node locations and effects, and in Section S2.1, we prove Theorem 3 in the main paper, which deals with the consistency of the estimates of the node locations and effects. The proof of Theorem 3 relies on proving consistency of the estimates of the global parameters, which we do in Section S2.2. Section S2.3 discusses the assumptions made in Theorem 3 in the main paper and demonstrates that several conventional distributions used in the literature satisfies these assumptions. Section S3 contains the proof of Theorem 4 in the main paper. Section S4 provides proofs of the other theorems in the main paper. Section S5 contains the proof of Theorem 5 and Section S6 contains the proof of Theorem 6. Sections S7 and S8 provide additional simulations. Section S9 provides simulations to verify the consistency of the claims made in Theorem 3. Section S10 contains additional lemmas and results we use in the supplementary materials.

In the proofs, we use C to refer to constants or sequences of constants that can change from line to line, but critically these constants never depend on the graph size n nor the number of nodes with trait k, n_k.

S1. Consistency of Beta-Model and SBM Parameters (Theorems 1 and 2)

We begin with the beta-model. Before providing specifics, we first introduce the main ideas of the proof of Theorem 1, which shows that the estimators, computed using just ARD, proposed in (1) are consistent for the parameters of the beta-model. To do this, we first recall that (1) proposes a fixed point estimator $\hat{\nu}_i$ that satisfies $\hat{\nu}_i(t+1) = \phi(\hat{\nu}_i(t))$ for some known function ϕ, which depends only on the degree sequence. They also propose a consistent estimator of the parameter β, which also only depends on the degree of the nodes. Since ARD allows us to recover the degree of nodes in the survey, we can then directly apply the results of (1) to conclude Theorem 1. Before getting to the proof of Theorem 1, we now re-state Theorem 3 of (1), which we use in our proof of Theorem 1.

Proposition **S1.1** (Theorem 3 of (1)). *The fixed point estimator, as described in equations 17-18 of (1), satisfies*

$$\max_{1\le i\le \hat{n}} |\hat{\nu}_i - \nu_i^\star| \le C\sqrt{\frac{\log(n)}{n}}$$

with probability $1 - O(1/n^2)$ *for some constant* $C > 0$. *In addition, we have that* $\hat{\beta} \xrightarrow{p} \beta$ *as* $n \to \infty$.

Proof of Theorem 1. In the case of mutually exclusive and exhaustive traits, $d_i = \sum_{k=1}^K y_{ik}$. Since the fixed point estimation procedure proposed in (1, 2) depends only on the degree of each node, which we are able to estimate with ARD, we can then apply Theorem 3 of (1) to conclude Theorem 1 of the main paper. Theorem 3 of (1) requires several conditions (Conditions 1, 2, 3, and 5 of (1)), which are all satisfied under the assumptions of Theorem 1 of the main paper. □

We now give a brief overview of the proof of Theorem 2. The intuition is that the the ARD responses $\tilde{y}_i = (y_{i1}/n_1, \ldots, y_{iK}/n_K)$ converge, by the weak law of large numbers, to $Z_i = (\tilde{P}_{i1}, \ldots, \tilde{P}_{iK})$ at an exponentially fast rate in n. See Figure S1 for an illustration of this fact. Therefore, two nodes in the same community will be classified together with probability going to 1, and since the by assumption the Z_i are distinct, two nodes in different communities will eventually be classified into different communities. We want to emphasize again the differences between the problem we are studying here and classic clustering problems or community detection problems. Compared to classic clustering problems, in which the distribution of data does not change as the sample size grows, the data we are analyzing here, y_{ik}/n_k, is converging to its expectation at an exponentially fast rate. Therefore, as our sample size grows, it becomes easier to correctly cluster the ARD responses and therefore to correctly classify nodes into the right communities. Second, compared to more standard community detection problems, we do not observe the graph but instead observe ARD about the nodes (3). This ARD, because it is a sample average, converges exponentially fast to its mean, which allows us to perform fast community detection.

Proof of Theorem 2. To begin, we pick a node randomly from V. Let c_i denote its community membership. For any j, since y_{jk}/n_k is a sum of (conditionally) independent random variables, by Hoeffding's inequality we have that $\mathbb{P}(|y_{jk}/n_k - p_{jk}| > \epsilon_n) \le a\exp(-a'\epsilon_n^2 n_k)$ for constants a and a'. To simplify notation here, suppose that we have groups of equal size, so that $n_k = n/K$, but this is not required for our analysis. By recalling that $\tilde{y}_i = (y_{i1}/n_1, \ldots, y_{iK}/n_K)$ is the normalized ARD response with mean $\tilde{p}_i = (\tilde{P}_{i1}, \ldots, \tilde{P}_{iK})$, we can conclude by a union bound that

$$\mathbb{P}(\max_{j:c_j=c_i} ||\tilde{y}_j - \tilde{p}_j|| > \epsilon_n) \le n \times a\exp(-a'\epsilon_n^2 n)\ .$$

for some constants a and a'. By taking $\epsilon_n^2 = \log(n)/n$, we see that $\mathbb{P}(\max_{j:c_j=c_i} \mathbf{1}\{\hat{c}_j \neq \hat{c}_i\} > 0) \le 1/n$. In addition, since $\Delta := \min_{c,c'} ||Z_c - Z_{c'}|| > 0$, which gives us well-separated clusters, and $\epsilon_n \to 0$, we have that $\mathbb{P}(\max_{j:c_j\neq c_i} \mathbf{1}\{\hat{c}_j \neq \hat{c}_i\}) \to 1$ for any j with $c_j \neq c_i$. By definition of the classification algorithm, we can conclude that $\mathbb{P}(\max_{j:c_j=c_i} \mathbf{1}\{\hat{c}_j \neq \hat{c}_i\} > 0) \le \mathbb{P}(\max_{j:c_j=c_i} ||\tilde{y}_j - \tilde{p}_j||)$.

Since the algorithm assigns nodes j that are within ϵ_n away from i into the same category, we see that the probability of any incorrect classification goes to zero for this community. The same argument applies to the second community, when looking at the set $V \setminus \hat{C}_i$. We then repeat this argument until all nodes are classified.

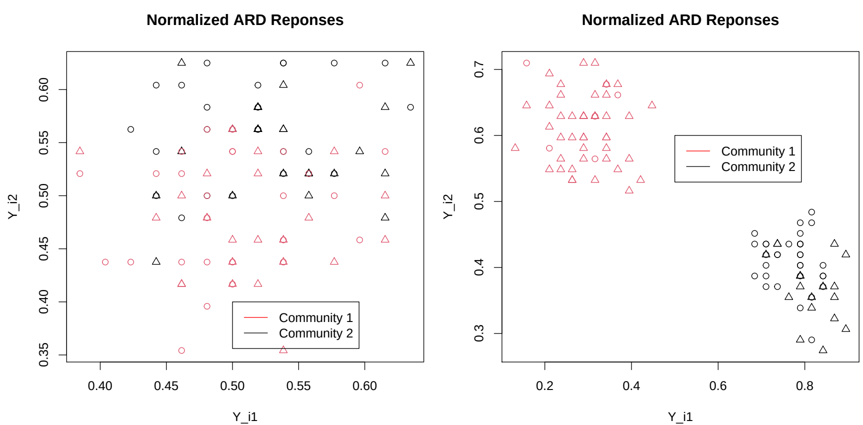


Fig. S1. Comparison of ARD responses in two different scenarios. On the left, we generate traits using the matrix $Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$. In this case, traits have no relationship with the community membership. In the left figure, we plot the normalized ARD responses, Here red indicates community 1, black indicates community 2, circles indicate trait 1, and triangles indicate trait 2. On the right, we repeat the simulation but using $Q = \begin{pmatrix} 7/10 & 3/10 \\ 1/10 & 9/10 \end{pmatrix}$. Here, there is a strong relationship between traits and community membership, and so K-means returns the correct clustering of the data.

Given a consistent estimate of the community membership vector, it follows from the weak law of large numbers that $\hat{Q}, \hat{P}$, and $\hat{\pi}$ are consistent for Q, P and π, where

$$\hat{Q}_{ck} = \frac{1}{m_c(n)} \sum_{i \in \hat{C}_c} \mathbf{1}\{t_i = k\}$$

$$\hat{P}_{cc'} = \begin{cases} \frac{1}{m_c(n)} \sum_{i \in \hat{C}_c} \sum_k \frac{y_{ik} \hat{\mathbb{P}}(c_j = c' | t_j = k)}{n_{c'}}, & c \neq c' \\ \frac{1}{m_c(n)} \sum_{i \in \hat{C}_c} \sum_k \frac{y_{ik} \hat{\mathbb{P}}(c_j = c' | t_j = k)}{n_{c'} - 1}, & c = c' \end{cases} .$$

where y_{ik} is the ARD response from node i about trait k, $\hat{\pi}_c = \frac{1}{m_c(n)} \sum_{i=1}^{n} \mathbf{1}\{\hat{c}_i = c\}$, and $m_c(n)$ is the number of nodes that we estimate to be in community c under the estimated community membership vector $\hat{\mathbf{c}}$. Here, recall that

$$\begin{aligned} \mathbb{P}(c_j = c' \mid t_j = k) &= \frac{\mathbb{P}(t_j = k, c_j = c')}{\mathbb{P}(t_j = k)} \\ &= \frac{\mathbb{P}(t_j = k \mid c_j = c')\mathbb{P}(c_j = c')}{\mathbb{P}(t_j = k)} \\ &= \frac{Q_{c'k} \times \pi_{c'}}{\mathbb{P}(t_j = k)} \end{aligned}$$

and we let $\hat{P}(c_j = c' | t_j = k)$ denote the estimate of this probability, computed by plugging in estimates for Q, π, and $P(t_j = k)$. □

S2. Consistency of Latent Space Model Parameters (Theorem 3)

We now define the estimates of the node locations and the node effects. In the estimates provided below, we assume that we have estimates of the global parameters, which we denote by $\eta^\star = (\mu_1^\star, \ldots, \mu_K^\star, \sigma_1^\star, \ldots, \sigma_K^\star, E\{\exp(\nu^\star)\})$. In Section S2.2, we provide estimates of $\eta^\star$ based on method-of-moment estimators.

Recall that the ARD data y_{ik} satisfies $y_{ik} \mid \nu_i^\star, z_i^\star, \eta^\star \sim \text{Binomial}(n_k, p_{ik})$ where n_k is the size of group k and p_{ik}, which we now define. With ARD data we do not observe any connections in the graph directly. It is possible, though unlikely as long as the sample size is small compared to the population size when using simple random sampling, that we might observe an alter of one of the surveyed respondents. That is, if person i reports knowing 5 people named Michael, one of those people named Michael might also be in the survey. Even in the unlikely event that this happens, we do not have access to this information through ARD since we do not observe any links. When considering the Binomial representation, therefore, we are making a statement not about the connections between any two individuals (which we do not observe) but instead about marginal connections between a person and a population. Respondent i is almost certainly more likely to know some members of the group k than others, but since ARD does not provide information on edges there is no way to specify that heterogeneity. Instead, we focus on an aggregate summary of the relationship between respondent i and members of group k which does not differ between members of the group because ARD, unlike the complete graph, does not contain sufficient data to do so. The power of our approach, however, is that, even under this limited information setting we still recover consistent estimates of model parameters.

Conditioned on node i's effect $\nu_i^\star$ and latent space location $z_i^\star$, the probability node i connects to an arbitrary node j in group k, written as is $\mathbb{P}(g_{ij} = 1 \mid \nu_i^\star, z_i^\star, \eta^\star) := p_{ik}$,

$$\begin{aligned} p_{ik} &= \int_V \int_Z \exp\{\nu_i^\star + \nu_j - d(z_i^\star, z_j)\} f_k(z_j) f_V(\nu_j) \; d\nu_j \; dz_j \\ &= \exp(\nu_i^\star) E\{\exp(\nu)\} \int_Z \exp\{-d(z_i^\star, z_j)\} f_k(z_j) \; dz_j \; . \end{aligned} \qquad \text{[S.1]}$$

Here, we use the notation $\nu_i^\star$ to refer to a fixed but unknown parameter of interest, whereas ν_j represents the variable that is integrated out. Note here we have used the property that $\exp(a + b) = \exp(a)\exp(b)$. By assuming the link function is exponential, we can easily separate the terms in the expression for $\mathbb{P}(g_{ij} = 1 \mid \nu_i, z_i, \eta)$. We believe we can extend these ideas to other link functions, as was done in (4), but we leave that to future work.

We now motivate and then formally describe these method-of-moment estimators (or equivalently, Z-estimators). Since the ARD is Binomial, we can estimate p_{ik} by equating p_{ik} with y_{ik}/n_k. This then allows us to solve for the parameters ν_i and z_i since p_{ik} depends on these two parameters (and η, which we can consistently estimate). In total, we create two systems of equations (one for the node locations and one for the fixed effects). This section assumes that we know the true parameters $\eta^\star$, but in Section S2.2 we show how to estimate the parameters $\eta^\star$.

We start with estimating the node locations. To do this, we note that the ratio $y_{ik}/y_{ik'}$ converges in probability, by the weak law of large numbers, to the ratio

$$E_{\sigma_k}[\exp\{-d(z_i, z)\}] / E_{\sigma_{k'}}[\exp\{-d(z_i, z')\}] \; ,$$

which depends only on the variances of the distributions of node locations $\sigma_1, \ldots, \sigma_K$ and the node location z_i, where we define the notation $E_\sigma[\exp\{-d(z, z_i)\}]$ to mean that the expectation is taken with respect to σ. Note that critically, in the ratio

$p_{ik}/p_{ik'}$, the terms involving the node effects and $E\{\exp(\nu)\}$, which are all unknown at this point, cancel out. This is the reason we look at the ratio of two ARD responses. Here we also make the simplifying assumption that $n_k = n_{k'}$, although the results do not change significantly if we remove this assumption. This suggests that we should take our estimate of the node location, denoted by $\hat{z}_i$, to be the value of z_i such that $y_{ik}/y_{ik'}$ is equal to the ratio $E_{\sigma_k}[\exp\{-d(z_i,z)\}]/E_{\sigma_{k'}}[\exp\{-d(z_i,z')\}]$.

More formally, we define the function $G_1 : \mathcal{M} \times (0,\infty)^2 \to \mathbb{R}$ by

$$G_1(z_i;\sigma_k,\sigma_{k'}) = \frac{E_{\sigma_k}[\exp\{-d(z_i,z)\}]}{E_{\sigma_{k'}}[\exp\{-d(z_i,z')\}]} \,. \qquad \text{[S.2]}$$

We drop the dependence on k and k' for simplicity and just write G_1 without any mention of k or k'. This function, when viewed as a function of z_i for a fixed $\sigma_k, \sigma_{k'}$, is not always invertible, but we can define a pseudo-inverse by $G_1^{-1}(x) = \{m \in \mathcal{M} : G_1(m) = x\}$. In the following calculations, we will take the inverse to be chosen in a fixed way from this set. We discuss this condition further and give examples in Section S2.3. Our estimate of the node location, $\hat{z}_i$, solves $\log\{G_1(\hat{z}_i;\hat{\sigma}_k,\hat{\sigma}_{k'})\} = \log(y_{ik}/n_k) - \log(y_{ik'}/n_{k'})$ for two arbitrary and distinct entries k, k'. In practice, the user selects the values of k and k'. The user can estimate a location using each pair of indices $k \neq k'$. Taking an average (or the Fréchet mean more generally) would improve the accuracy of the resulting estimate. Note that the log transformation simplifies the analysis of this estimator and allows us to use a proof technique that is similar to the one used to prove Theorem 1.3 in (2) or Theorem 3 in (1).

We now motivate our estimator of the the node effects. The idea is that ARD is a Binomial random variable and thus we can equate the probability of an edge between node i and nodes in group k (which depends on the node effect and the node location, which we have already estimated above) with the observed number of edges. We then solve for the node effect. To state this estimator more formally, define the function

$$G_2(\nu_i, z_i) = E\{\exp(\nu)\}\exp(\nu_i)E[\exp\{-d(z_i,z)\}] \,,$$

where here $z \sim F(\mu_k, \sigma_k^2)$. Since y_{ik}/n_k converges in probability to $G_2(z_i^\star, \nu_i^\star)$, this motivates the following estimator

$$\hat{\nu}_i = \log\left(\frac{y_{ik}}{n_k}\right) - \log(E[\exp\{-d(\hat{z}_i,z)\}]) - \log[\hat{E}\{\exp(\nu)\}] \,. \qquad \text{[S.3]}$$

where $z \sim F(\hat{\mu}_k, \hat{\sigma}_k)$ and the term $\log[\hat{E}\{\exp(\nu)\}]$ is the estimate of $\log[E\{\exp(\nu)\}]$ computed using $\hat{\eta}$. Again, as in the case of the node locations, the user can select the group index k used in computing $\hat{\nu}_i$. As in the case of the node location, we can compute $\hat{\nu}_i$ for all group indices k and their average will be an improved estimate of $\nu_i^\star$.

In the next section, we prove Theorem 3 in the main paper, which deals with showing that estimates of the node locations and node effects are consistent and satisfy a convergence rate of $\sqrt{3\log(\tilde{n})/2\tilde{n}}$ with probability at least $1 - O(m/\tilde{n}^3)$, where $\tilde{n} = n/K$ and K is assumed to be fixed. Our proof of Theorem 3 is based on two separate lemmas: Lemma S2.2 proves the claimed convergence result for the node locations, and Lemma S2.3 proves the claimed convergence result for the node effects.

To begin with some notation, the estimates of the node locations and the node effects depend on the group parameters, which we denote by η. We let $\hat{z}_i(\eta)$ denote the estimate of $z_i^\star$ that is computed using the known and true η, and we let $\hat{z}_i(\hat{\eta})$ denote the estimate based upon the plug-in estimate $\hat{\eta}$, which we define formally in Section S2.2.

S2.1. Proof of Theorem 3. We now provide a proof of Theorem 3 in the main text. For clarity, we repeat the statement of the proof here along with the necessary assumptions. The proof relies on consistent estimates of the global parameters. For ease of exposition, we have moved the derivation of these estimates to the subsequent section. We prove the result by constructing a series of Lemmas that, when combined, yield the desired result. We begin by restating the necessary assumptions. Additional discussion of the assumptions, including verification that they hold with distributional assumptions commonly used in practice is in Section S2.3. Note that in the main part of the paper, the following four assumptions are labeled as Assumptions 2-5.

ASSUMPTION **S2.1.** *For each k, μ_k is in a compact subset of $\mathcal{M}^p(\kappa)$ and σ_k is in a compact subset of $(0,\infty)$.*

ASSUMPTION **S2.2.** *The node effects $\nu_i^\star \overset{iid}{\sim} H$ satisfy $E\{\exp(\nu_i^\star)\} < \infty$.*

ASSUMPTION **S2.3.** *The distribution F is a symmetric distribution on $\mathcal{M}^p(\kappa)$ that is completely characterized by its mean and variance and satisfies the following two conditions. The function $z_i \mapsto E_k[\exp\{-d(z_i,z)\}]$ is Lipschitz for every $k \in \{1,\ldots,K\}$ and $z_i \mapsto E_k[\exp\{-d(z_i,z)\}]/E_{k'}[\exp\{-d(z_i,z')\}]$ has a pseudo-inverse that is Lipschitz.*

ASSUMPTION **S2.4.** *Define $F_1 : (z_i,\sigma_k,\sigma_{k'}) \mapsto E_k[\exp\{-d(z_i,z)\}]/E_{k'}[\exp\{-d(z_i,z')\}]$. The inverse function F_1^{-1} is continuous in σ and for every k, k', ℓ, and ℓ', the following two functions are Lipschitz:*

$$\eta \mapsto \frac{E_{kk'}[\exp\{-d(z,z')\}]}{E_{\ell\ell'}[\exp\{-d(z,z')\}]}, \quad \eta \mapsto \frac{E_{kk'}[\{\exp(-d(z,z')\}]^2}{E_{\ell\ell'}[\{\exp(-d(z,z')\}]^2} \,.$$

Under the four assumptions above, we now restate Theorem 3 in the main paper.

Theorem 1. *Suppose Assumptions S2.1, S2.2, S2.3, and S2.4 hold. The estimators $\hat{z}_i$ and $\hat{\nu}_i$ and $\hat{\eta}$ are consistent for $z_i^\star, \nu_i^\star$, and $\eta^\star$ as $m, n \to \infty$, up to isometry on $\mathcal{M}^p(\kappa)$ and satisfy*

$$\max_{1 \le i \le m(n)} d_{\mathcal{M}^p(\kappa)}(\hat{z}_i, z_i^\star) \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}},$$

$$\max_{1 \le i \le m(n)} |\hat{\nu}_i - \nu_i^\star| \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}},$$

with probability $1 - O(m/\tilde{n}^3)$.

Proof of Theorem 3 in the main paper. For readability, we split up the proof of Theorem 3 in the main paper into several lemmas. Theorem 3 claims a concentration inequality for the estimates of the node locations and node effects using the plug-in estimate $\hat{\eta}$ of the global parameters. We prove this result for the node locations (Lemma S2.2) and for the node effects (Lemma S2.3) separately. These two lemmas require us to first prove the consistency (without a rate) on the estimates of node locations and effects, which we do in Lemma S2.1. The proofs of Lemmas S2.2 and S2.3 are based on Lemmas S2.4 and S2.5, which prove the concentration inequalities using the true and unknown group parameter η. Combining the arguments in these lemmas proves the desired result.

□

Our proof of Theorem 3 starts with the following lemma, which states the estimates that maximize the pseudo-likelihood of the ARD are consistent as $m, n \to \infty$. We use this result later on to prove Theorem 3. We would like to emphasize that maximizing the pseudo-likelihood, which we do in Section S10, is equivalent to a method-of-moments estimator in this case.

LEMMA **S2.1.** *Let the assumptions from Theorem 3 of the main paper hold. Suppose that we have consistent estimates of the group parameters η, denoted by $\hat{\eta}$. Now suppose that $(\hat{\nu}_{1:m}, \hat{z}_{1:m})$ are the Z-estimators of the node effects and locations described in Section S2. Then, $(\hat{\nu}_{1:m}, \hat{z}_{1:m})$ are consistent for $\nu^\star_{[1:m]}$ and $z^\star_{[1:m]}$ as $m, n \to \infty$, up to an isometry on $\mathcal{M}^p(\kappa)$.*

For readability, we have moved the proof of Lemma S2.1 to Section S10. The main idea of the proof follows the standard M-estimator consistency steps: showing a well-separated extremum and a uniform law of large numbers (5).

LEMMA **S2.2.** *With probability at least $1 - O(m/\tilde{n}^3)$, the following inequality holds up to isometry on $\mathcal{M}^p(\kappa)$.*

$$\max_{1 \le i \le m(n)} d_{\mathcal{M}}(\hat{z}_i(\hat{\eta}), z_i^\star) \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}}\,.$$

Proof. By the triangle inequality,

$$d_{\mathcal{M}}(\hat{z}_i(\hat{\eta}), z_i^\star) \le d_{\mathcal{M}}(\hat{z}_i(\hat{\eta}), \hat{z}_i(\eta)) + d_{\mathcal{M}}(\hat{z}_i(\eta), z_i^\star). \qquad \text{[S.4]}$$

We have two terms in the triangle inequality. We will only have to focus on the second one, because that will dominate the rate as we will soon show. We calculate that one below. The first one has an extremely fast rate as it tends to zero. This can be seen in a straightforward manner from using a Taylor expansion of the estimating equation in the usual way, because the estimating equation consists of an average taken over all pairs of groups and all pairs of potential links across every pair of group which gives order $O_P(1/\sqrt{K^2mn})$, where again m is the size of the ARD sample. We will show later that this rate is much faster than the rate for the second term in the inequality, which means this term can be ignored when proving the rate of convergence on the term $d_{\mathcal{M}}(\hat{z}_i(\hat{\eta}), z_i^\star)$.

We now study the second term in the triangle inequality above. Now, using the definition of $\hat{z}_i(\eta)$ as $\hat{z}_i = G_1^{-1}(a; \hat{\eta})$, we write

$$d_{\mathcal{M}}(\hat{z}_i(\hat{\eta}), \hat{z}_i(\eta)) = d_{\mathcal{M}}(G_1^{-1}(a; \hat{\eta}), G_1^{-1}(a; \hat{\eta}))$$

where $a = \log(y_{ik}/n_k) - \log(y_{ik'}/n_{k'})$.

Supposing that $G_1^{-1}(a; \sigma)$ is continuous in σ, which we assume in Theorem 3 in the main paper, we combine Lemma S2.6 with the continuous mapping theorem to show that $d_{\mathcal{M}}(\hat{z}_i(\hat{\eta}), \hat{z}_i(\eta))$ converges to zero in probability. All we need to do now is show that the second term in (S.4) satisfies the claimed concentration inequality. By Lemma S2.4, which we state below, with probability at least $1 - O(1/n_k^3)$,

$$d_{\mathcal{M}}(\hat{z}_i(\eta), z_i^\star) \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}}\,,$$

up to isometry on $\mathcal{M}$. By a union bound, and by recalling (S.4), we conclude that with probability at least $1 - O(m/\tilde{n}^3)$:

$$\max_{1 \le i \le m(n)} d_{\mathcal{M}}(\hat{z}_i(\hat{\eta}), z_i^\star) \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}}$$

up to isometry on $\mathcal{M}$. □

The next lemma shows that the estimate of ν_i, based on the plug-in estimate $\hat{\eta}$, satisfies a similar concentration inequality.

LEMMA **S2.3.** *The estimator $\hat{\nu}_i$ from (S.3) satisfies the following: With probability $1 - O(m/\tilde{n}^3)$,*

$$\max_{1 \le i \le m(n)} |\hat{\nu}_i(\hat{\eta}) - \nu_i^\star| \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}} \,.$$

Proof. The proof follows the same argument that we used in the proof of Lemma S2.2. Since $\hat{\eta}$ is consistent for η, the second term in the definition of $\hat{\nu}_i$ can be ignored when proving the desired concentration inequality (again, this argument was used in the proof of Theorem 3 in (1)). It therefore suffices to just argue that the term $\log(y_{ik}/n_k)$ satisfies the claimed concentration inequality. We can prove this inequality by Hoeffding's inequality. See Lemma S2.5, which proves this formally. Taking a union bound over all $i = 1, \dots, m(n)$ to proves the desired result. □

In the case where $d(z_i, z_j) = 0$ (only node effects determine connection propensity) and $m = n$ (meaning that we observe the entire graph and not just the ARD), then Theorem 3 of the main paper simplifies to Theorem 3.3 of (2).

LEMMA **S2.4.** *With probability at least $1 - O(m/\tilde{n}^3)$, the following inequality holds:*

$$\max_{1 \le i \le m(n)} d_{\mathcal{M}}(\hat{z}_i(\eta), z_i^\star) \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}} \,.$$

The proof is based on similar ideas found in (1, 2). The intuition behind the proof is as follows. The estimator $\hat{z}_i(\eta)$ is based on the ARD $y_{ik}/n_k = 1/n_k \sum_{j \in G_k} g_{ij}$, which converges exponentially fast to p_{ik} by Hoeffding's inequality. This insight allows us to conclude the uniform control over the error in $\hat{z}_i(\eta)$.

Proof. To begin, we recall that the estimator is $\hat{z}_i = G_1^{-1}(y_{ik}/n_k; \eta)$. This function will not be invertible, but we can choose a representative from the set of $\{x : G_1(x; \eta) = y_{ik}/n_k\}$. Any choice will lead to the right answer, up to isometry. Note also that because of properties of $\mathcal{M}^p(\kappa)$, it is locally Euclidean. See (4) and its references for a more complete description of this point. Since $\hat{z}_i(\hat{\eta})$ converges to $z_i(\eta)$, up to isometry, we therefore only need to prove the argument for the Euclidean case (this follows from Lemma S2.1). The extension to the spherical and hyperbolic geometries follows since there is a neighborhood around z_i in which the distances are approximately Euclidean distances, and thus the Euclidean arguments apply here too.

Since

$$a = \log(y_{ik}/n_k) - \log(y_{ik'}/n_{k'})$$

converges in probability, as $n \to \infty$, to $G_1(z_i)$, this motivates our estimate of z_i. We set $\hat{z}_i = G_1^{-1}(a)$. See Section S2.3 for a discussion on this inverse function. Since $G_1^{-1}\{\log(p_{ik}) - \log(p_{ik'})\} = z_i^\star$,

$$\begin{aligned} \|\hat{z}_i(\eta) - z_i^\star\| &= \left\| G_1^{-1}(a) - G_1^{-1}\{\log(p_{ik}) - \log(p_{ik'})\} \right\| \\ &\le C_n |\log(y_{ik}/n_k) - \log(y_{ik'}/n_{k'}) - \log(p_{ik}) - \log(p_{ik'})| \\ &\le \tilde{C}_n \{|y_{ik}/n_k - p_{ik}| + |y_{ik'}/n_{k'} - p_{ik'}|\} \,. \end{aligned}$$

where C_n and $\tilde{C}_n$ are sequences of constants. We know that $\tilde{C}_n$ is on the order $n_k = O(n)$ when K is fixed (which we assume), since $x \mapsto \log(x)$ is Lipschitz on any interval $[a', b']$ with Lipschitz constant $1/a'$. In our case, with probability going to 1, $y_{ik} \ge 1$ and so $y_{ik}/n_k \ge 1/n_k$ and thus we can take $1/(1/n_k) = n_k$ to be the Lipschitz constant. We thus conclude that

$$\mathbb{P}(\|\hat{z}_i(\eta) - z_i^\star\| > \epsilon) \le \mathbb{P}\left(\left| \frac{y_{ik}}{n_k} - p_{ik} \right| > \epsilon/\tilde{C}_n \right) + \mathbb{P}\left(\left| \frac{y_{ik'}}{n_{k'}} - p_{ik'} \right| > \epsilon/\tilde{C}_n \right) \,. \quad \text{[S.5]}$$

We now show that both terms on the right hand side converge to zero exponentially fast. Since y_{ik} is a sum of independent Bernoulli random variables, each with expectation p_{ik}, by Hoeffding's inequality (6),

$$\mathbb{P}\left(\left| \frac{y_{ik}}{n_k} - p_{ik} \right| > \epsilon/\tilde{C}_n \right) \le 2\exp\left(-2\frac{\epsilon^2 n_k}{\tilde{C}_n^2} \right) \,.$$

Set $\epsilon^2 = \frac{3}{2} n_k^{-1} \tilde{C}_n^2 \log(n_k) = O(\frac{3}{2} n_k^{-1} n_k^2 \log(n_k))$. Then,

$$\mathbb{P}\left(\left| \frac{y_{ik}}{n_k} - p_{ik} \right| > \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}} \right) \le 2\exp\{-3\log(n_k)\} = 2/n_k^3 \,.$$

Similarly, $\mathbb{P}\left(\left| \frac{y_{ik'}}{n_{k'}} - p_{ik'} \right| > \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}} \right) \le 2/n_k^3$. Putting this together, and recalling (S.5), we see that

$$\mathbb{P}\left(\|\hat{z}_i(\eta) - z_i^\star\| > \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}} \right) \le 4/n_k^3 \,.$$

By a union bound, with probability at least $1 - 4m/n_k^3$,

$$\max_{1 \le i \le m} \|\hat{z}_i(\eta) - z_i^\star\| < \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}} \,.$$

□

In the following lemma, we prove that the estimate $\hat{\nu}_i$ satisfies a similar type of concentration inequality. The proof is identical to the one given above, so we omit the details.

LEMMA **S2.5.** *If each z_i is known, and the global parameter η is known, the estimator $\hat{\nu}_i$ defined in (S.3) satisfies the following: With probability at least $1 - O(m/\tilde{n}^3)$,*

$$\max_{1 \le i \le m(n)} |\hat{\nu}_i(\eta) - \nu_i| \le \sqrt{\frac{3\log(\tilde{n})}{2\tilde{n}}} .$$

S2.2. Estimating Global Parameters in Latent Space Model. In this section, we provide estimates of the model parameters η. Our discussion comes in three parts. We first show how to estimate the within-group variance terms. To estimate the within-group variances, we equate the ARD responses of people in a group k to other nodes in the same group k with the probability that an arbitrary edge exists between nodes in group k. Since this probability depends on only the within-group variance, as all nodes from a given group are distributed about the same group center, we can therefore estimate the group variance in this way.

To formally define our estimator, fix two groups G_k and $G_{k'}$. The probability that an arbitrary node in group k connects to other nodes in group k is equal to, after integrating out all the parameters, $E\{\exp(\nu)\}^2 E_{kk}[\exp\{-d(z,z')\}]$, where z, z' are independent and $z, z' \sim F(\mu_k^\star, \sigma_k^\star)$. Note critically that this does not upon the mean parameter $\mu_k^\star$.

We let $m_k(n)$ be the number of nodes we sample that belong to group k. We define the quantity

$$t_{kk'} = \frac{1}{m_k(n)} \sum_{i \in G_k} \frac{y_{ik'}}{n_{k'}} . \qquad \text{[S.6]}$$

Then, for large n (which implies that $|G_k| = n_k$ and $m_k(n)$ is large too), the ratio $t_k/t_{k'}$ converges in probability to

$$\frac{E\{\exp(\nu)\}^2 E_{kk}[\exp\{-d(z,z')\}]}{E\{\exp(\nu)\}^2 E_{k'k'}[\exp\{-d(z,z')\}]} = \frac{E_{kk}[\exp\{-d(z,z')\}]}{E_{k'k'}[\exp\{-d(z,z')\}]} . \qquad \text{[S.7]}$$

which depends again on just the unknown variance terms $\sigma_k^\star$ and $\sigma_{k'}^\star$. In other words, by looking at the ratio $t_k/t_{k'}$, the term $E(\exp(\nu))^2$, which we have not yet estimated and do not know in practice, cancels. So this ratio depends only on the unknown variance vector $(\sigma_1^\star, \dots, \sigma_\star^2)$. Motivated by this description, we define an estimator $\hat{\sigma}^2(n) = \{\hat{\sigma}_1^2(n), \dots, \hat{\sigma}_K^2(n)\}$ as the root of the following system of equations

$$\frac{t_{kk}}{t_{k'k'}} = \frac{E_{kk}[\exp\{-d(z,z')\}]}{E_{k'k'}[\exp\{-d(z,z')\}]} . \qquad \text{[S.8]}$$

If K is large enough to ensure the above solution has a unique zero in the limit as $m, n \to \infty$, this estimator is consistent for the true $(\sigma_1^\star, \dots, \sigma_K^\star)$.

LEMMA **S2.6.** *The estimator $\hat{\sigma}^2(n) = \{\hat{\sigma}_1^2(n), \dots, \hat{\sigma}_K^2(n)\}$ that is the root of the system from (S.8) is consistent as $n \to \infty$.*

Proof. We first sketch an outline of our argument. We will define a sequence of random functions $\hat{H}_n$ such that $\lim_n E\{\hat{H}_n(\sigma^2)\} = 0$ only at the true $\sigma^\star$. This sequence of functions $\hat{H}_n$ is defined such that the estimator from the lemma minimizes this expression. Thus, to show consistency of the estimator, we can simply verify the two conditions from Theorem 5.7 of (5), which for completeness we give in Section S10. At a high level, Condition 1 requires that H have a well-separated zero, and Condition 2 requires that $\hat{H}_n$ converge uniformly to H. Once we verify these two conditions, we can then conclude from Theorem 5.7 of (5) the desired consistency result.

By recalling the definition of t_k in (S.6), we define the sequence of random functions $\hat{H}_n : (0, \infty)^K \to [0, \infty)$ by

$$\hat{H}_n(\sigma^2) = \sum_{k=1}^K \sum_{k'=1}^K \left\{ \frac{t_{kk}}{t_{k'k'}} - \frac{E_{kk}[\exp\{-d(z,z')\}]}{E_{k'k'}[\exp\{-d(z,z')\}]} \right\}^2 .$$

We then define $H_n(\sigma^2) = E\{\hat{H}_n(\sigma^2)\}$ and $H(\sigma^2) = \lim_{n\to\infty} H_n(\sigma^2)$. By (S.7) and using the weak law of large numbers, combined with the continuous mapping theorem, it is clear that H evaluated at the true σ^2 is zero. For sufficiently large K, this zero is unique, by using the same argument that we give in Lemma S10.3 or by using Theorem 3 of (7). So Condition S10.1 is satisfied.

We now prove Condition S10.2. Recall that our goal is to show that

$$\sup_{\sigma^2 \in S} |\hat{H}_n(\sigma^2) - H(\sigma^2)| \xrightarrow{p} 0$$

It suffices to show that $\sup_{\sigma^2 \in S} |\hat{H}_n(\sigma^2) - H_n(\sigma^2)| = o_P(1)$, because H_n converges uniformly to H deterministically and hence also in probability. To show *this* uniform law of large numbers, we will use Corollary 2.1 of (8). For completeness, we provide this corollary in Section S10. The pointwise convergence is automatically satisfied, by recalling (S.7). We now fix a k, k' and expand inside the double sum in the expression for $\hat{H}_n$ as

$$\frac{t_{kk}}{t_{k'k'}} - 2\frac{t_{kk}}{t_{k'k'}} \frac{E_{kk}[\exp\{-d(z,z')\}]}{E_{k'k'}[\exp\{-d(z,z')\}]} + \frac{E_{kk}[\exp\{-d(z,z')\}]^2}{E_{kk}[\exp\{-d(z,z')\}]^2} .$$

By comparing the terms inside the expression $|\hat{H}_n(\sigma^2) - \hat{H}_n(\tilde{\sigma}^2)|$, we see that there are just two terms to consider. To show the Lipschitz condition required to use Corollary 2.1 of (8), let $\sigma, \tilde{\sigma} \in S \subseteq (0,\infty)^K$. To simplify the notation, we let $E_{kk}[\exp\{-d(z,z')\}]$ denote the expectation using the variance vector σ and $\tilde{E}_{kk}[\exp\{-d(z,z')\}]$ to denote the expectation using the variance $\tilde{\sigma}$.

By assumption, the first term satisfies

$$2\frac{t_{kk}}{t_{k'k'}}\left|\frac{E_{kk}[\exp\{-d(z,z')\}]}{E_{k'k'}[\exp\{-d(z,z')\}]} - \frac{\tilde{E}_{kk}[\exp\{-d(z,z')\}]}{\tilde{E}_{k'k'}[\exp\{-d(z,z')\}]}\right| \leq C\frac{t_{kk}}{t_{k'k'}}||\sigma^2 - \tilde{\sigma}^2|| \ ,$$

where C is a constant. By assumption, the second term satisfies a similar Lipschitz condition:

$$\left|\frac{E_{kk}[\exp\{-d(z,z')\}]^2}{E_{k'k'}[\exp\{-d(z,z')\}]^2} - \frac{\tilde{E}_{kk}[\exp\{-d(z,z')\}]^2}{\tilde{E}_{k'k'}[\exp\{-d(z,z')\}]^2}\right| \leq C'||\sigma^2 - \tilde{\sigma}^2|| \ ,$$

where C' is a constant. Putting this all together, we see that

$$|\hat{H}_n(\sigma^2) - \hat{H}_n(\tilde{\sigma}^2)| \leq \sum_{k,k'}(C\frac{t_{kk}}{t_{k'k'}} + C')||\sigma^2 - \tilde{\sigma}^2|| \ .$$

Since $\sum_{k,k'} E(Ct_{kk}/t_{k'k'} + C') = O(1)$, we conclude by Corollary 2.1 of (8) that Condition 2 holds. By Theorem 5.7 of (5), we conclude the consistency claim in the theorem.

S2.2.1. Estimating Group Means. In this section, we show how to use the consistent estimates of the within-group variances $\sigma_1^\star, \ldots, \sigma_K^\star$ to estimate the group mean parameters. Motivated by the same approach we used to prove consistency of $\sigma_1^\star, \ldots, \sigma_K^\star$, consider now four group centers. The probability that nodes in the first two groups, say k and k' connect, divided by the probability that nodes in the last two groups, say ℓ and ℓ', connect is

$$\frac{E\{\exp(\nu)\}^2 E_{kk'}[\exp\{-d(z,z')\}]}{E\{\exp(\nu)\}^2 E_{\ell\ell'}[\exp\{-d(z,z')\}]} = \frac{E_{kk'}[\exp\{-d(z,z')\}]}{E_{\ell\ell'}[\exp\{-d(z,z')\}]} \ .$$

Having estimated the within-group variances terms, and noting that $t_{kk'}/t_{\ell\ell'}$ estimates the probability above, we can estimate the terms $\mu_1^\star, \ldots, \mu_K^\star$ by solving the following system of equations: for every 4-tuple (k, k', ℓ, ℓ') with distinct entries,

$$\frac{t_{kk'}}{t_{\ell\ell'}} = \frac{E_{kk'}[\exp\{-d(z,z')\}]}{E_{\ell\ell'}[\exp\{-d(z,z')\}]} \ . \qquad \text{[S.9]}$$

The following lemma shows that this estimator is consistent as $n \to \infty$.

LEMMA **S2.7.** *Let $\hat{\mu}_1(n), \ldots, \hat{\mu}_K(n)$ be a root of the system in (S.9). This estimator is consistent as $n \to \infty$, up to an isometry on $\mathcal{M}$.*

Proof. The proof is nearly identical to the one given for Lemma S2.6, so we only sketch the argument. We define the sequence of random functions

$$\hat{H}_n(\mu) = \sum_{k,k,\ell,\ell'}\left\{\frac{t_{kk'}}{t_{\ell\ell'}} - \frac{E_{kk'}[\exp\{-d(z,z')\}]}{E_{\ell\ell'}[\exp\{-d(z,z')\}]}\right\}^2$$

We also define $H_n(\mu) = E\{\hat{H}_n(\mu)\}$ and $H(\mu) = \lim_{n\to} H_n$. At the true $\mu^\star$ parameter, $H(\mu^\star) = 0$ for sufficiently large K. For sufficiently large K, this is the only zero, up to an isometry on $\mathcal{M}$. (Again, by using the same argument that we give in Lemma S10.3 or by using Theorem 3 of (7).) Thus, Condition 1 is satisfied. To show Condition 2, we use the same argument as we give in the proof of Lemma S2.6. By assumption, we know that Condition 2 holds. Thus, by Theorem 5.7 of (5), we can conclude the desired consistency result. ☐

S2.2.2. Estimating Node Effect Expectation. In the previous two sections, we have shown how to obtain consistent estimates of the within-group variances and the group means. In this section, we show how to estimate the term $\tau = E[\{\exp(\nu)\}^2]$. The probability that any node in group k connects with any node in group k' is, after integrating out all parameters,

$$E[\{\exp(\nu)\}^2]E_{kk'}[\exp\{-d(z,z')\}] \ , \qquad \text{[S.10]}$$

where $z \sim F(\mu_k^\star, \sigma_k^\star)$ and $z' \sim F(\mu_{k'}^\star, \sigma_{k'}^\star)$. By drawing $\hat{z} \sim F(\hat{\mu}_k, \hat{\sigma}_k)$ independently of $\hat{z}' \sim F(\hat{\mu}_{k'}, \hat{\sigma}_k)$, we can use $E_{kk'}[\exp\{-d(\hat{z},\hat{z}')\}]$ to estimate the quantity $E_{kk'}[\exp\{-d(z,z')\}]$. Since

$$t_{kk'} = \frac{1}{n_k}\sum_{i\in G_k}\frac{y_{ik'}}{n_{k'}}$$

converges in probability to the expression in (S.10), we can estimate $E[\{\exp(\nu)\}^2]$ by

$$\hat{\tau} = \frac{t_{kk'}}{E_{kk'}[\exp\{-d(\hat{z},\hat{z}')\}]}.$$

where $\hat{z} \sim F(\hat{\mu}_k, \hat{\sigma}_k)$ independently of $\hat{z}' \sim F(\hat{\mu}_{k'}, \hat{\sigma}_k)$. By the continuous mapping theorem and by recalling (S.10), we can consistently estimate τ. ☐

S2.3. Discussion of Assumptions for Theorem 3.

In this section we discuss two of the assumptions made in the main paper and discuss when these hold.

The p-dimensional normal distribution in $\mathbb{R}^p$ and the von-Mises Fisher distribution on the p-sphere are two models commonly used in the literature. We now argue that these two model satisfy this assumption. Recall that the term in question, in the case of a p-dimensional Gaussian distribution, is

$$z_i \mapsto \int_{\mathbb{R}^p} \exp(-||z_i - z||)f(z)dz \ ,$$

where f here is the pdf of the p-dimensional Gaussian distribution. Note that $z \mapsto d(z_i, z)$ is Lipschitz, and $x \mapsto \exp(-x)$ is Lipschitz over $[0, \infty)$, and thus since $\exp(-x)$ is bounded by 1 on $(0, \infty)$, we conclude that $z_i \mapsto \exp\{-d(z_i, z)\}$ is Lipschitz. Because the integral of a Lipschitz function is again Lipschitz, we conclude that the assumption holds.

We now look at the assumption that the inverse of the function $z_i \mapsto G_1(z)$ is invertible, where G_1 is defined in (S.2). To begin the discussion, recall the simulation exercise in Figure S7. There are two group centers at $(2, 2)$ and $(-2, -2)$ in $\mathbb{R}^2$. The point we wish to estimate is at (0, 0), so the distance between each group center and this point is $2\sqrt{2}$. There is a unique point in $\mathbb{R}^2$ that satisfies this constraint. However, consider the following two examples.

EXAMPLE **S2.1.** *Consider two group centers at $(2, 2)$ and $(-2, -2)$ in $\mathbb{R}^2$. Suppose the point of interest z_i is 2 unit away from the first point and 2 away from the second point. Then, the points $(2, -2)$ and $(-2, 2)$ will both solve the expression $F(z) = \log(p_{ik}) - \log(p_{ik'})$, where p_{ik} depends on the distance between z_i and the group centers.*

EXAMPLE **S2.2.** *Now let $\mathcal{M}^p(\kappa) = S^1(1)$, the circle with radius 1. Set two group centers at $(0, 1)$ and $(-1, 0)$ and suppose that the point of interest is $\pi/2$ away from the first group center and $3\pi/2$ away from the second group center. Then there are two points at $(0, 1)$ and $(0, -1)$ that solve the expression $F(z) = \log(p_{ik}) - \log(p_{ik'})$, where p_{ik} depends on the distance between z_i and the group centers.*

The discussion above highlights the fact that the mapping $z \mapsto G_1(z)$ might not be invertible. We therefore suggest that the user select a representative element of the pseudo-inverse (hence our language in the main part of the paper).

We now turn to discussing Assumption S2.4. We show that under mild distributional assumptions, the function $\sigma \mapsto \frac{E_{kk'}[\exp\{-d(z,z')\}]}{E_{\ell\ell'}[\exp\{-d(z,z')\}]}$ is Lipschitz. The discussion of the function $\mu \mapsto \frac{E_{kk'}[\exp\{-d(z,z')\}]}{E_{\ell\ell'}[\exp\{-d(z,z')\}]}$ is very similar. Suppose first that the function $\sigma_k \mapsto E[\exp\{-d(z_i, z)\}]$ is Lipschitz. Then, suppose that $g : (\sigma_k, \sigma_{k'}) \mapsto E(\exp\{-d(z_i, z)\})/E(\exp\{-d(z_i, z')\})$ is differentiable. It then has a gradient $\nabla g = (\nabla_k g, \nabla_{k'} g)$, where

$$\nabla_k g = \frac{\partial g}{\partial \sigma_k} = \frac{d}{d\sigma_k} E[\exp\{-d(z_i, z)\}]]/E[\exp\{-d(z_i, z')\}]]$$

Supposing that $E[\exp\{-d(z_i, z)\}]]$ is bounded away from zero, then this partial derivative is bounded because we assumed that the function $\sigma_k \mapsto E[\exp\{-d(z_i, z)\}]]$ is Lipschitz. The other partial derivative is given by

$$\frac{\partial g}{\partial \sigma_{k'}} = E[\exp\{-d(z_i, z)\}]]/\frac{d}{d\sigma_{k'}} E[\exp\{-d(z_i, z')\}]]$$

Supposing that the function $\sigma_{k'} \mapsto E[\exp\{-d(z_i, z')\}]]$ has a derivative that is bounded away from zero, we can thus conclude that g is Lipschitz since each of its partial derivatives is bounded.

We now verify when the function $\sigma_k \mapsto E[\exp\{-d(z_i, z)\}]$ is Lipschitz. This function is given by

$$\sigma_k \mapsto \int_{\mathcal{M}} \exp\{-d(z_i, z)\} f_k(\mu_k, \sigma_k)dz \ .$$

Supposing that $\sigma_k \mapsto f_k(\mu_k, \sigma_k)$ is Lipschitz, then we can use the Leibnitz rule (which allows us to pass the derivative inside the integral) to conclude that the function $\sigma_k \mapsto E[\exp\{-d(z_i, z)\}]$ is Lipschitz. By explicitly calculating the derivative of this expression in the cause of a Gaussian distribution, we see that $\sigma_k \mapsto f_k(\mu_k, \sigma_k)$ is Lipschitz. Since by assumption, each σ_k is in a compact (and hence bounded subset of $(0, \infty)$), we can conclude that for each z_i, $\frac{\partial g}{\partial \sigma_k}$ is bounded. To show this, we need to show that $\frac{d}{d\sigma_{k'}} E[\exp\{-d(z_i, z')\}]]$ is bounded away from zero, which for a fixed z_i is true because the σ_k are by assumption in a compact subset of $(0, \infty)$. A similar argument applies to the function $\eta \mapsto \frac{E_{kk'}[\exp\{-d(z,z')\}]^2}{E_{\ell\ell'}[\exp\{-d(z,z')\}]^2}$.

S3. Consistency of plug-in estimator $E\{S_i(g_n) \mid \hat{\theta}_n(\mathbf{y})\}$ for $S_i(g_n^\star)$ (Theorem 4)

Proof of Theorem 4. By the triangle inequality,

$$|E\{S_i(g_n) \mid \hat{\theta}_n(\mathbf{y})\} - S_i(g_n^\star)| \le |E\{S_i(g_n) \mid \hat{\theta}_n(\mathbf{y})\} - E\{S_i(g_n) \mid \theta_n\}| + |E\{S_i(g_n) \mid \theta_n\} - S_i(g_n^\star)| \ .$$

By Condition 2 of Theorem 4, $|E\{S_i(g_n) \mid \theta_n\} - S_i(g_n^\star)| = o_P(1)$. We now analyze the other term. Under Condition 3, the function $\theta_n \mapsto E\{S_i(g_n) \mid \theta_n\}$ is differentiable, so by the mean value theorem, there exists a sequence of intermediate values $\bar{\theta}_n$ such that

$$E\{S_i(g_n) \mid \hat{\theta}_n(\mathbf{y})\} = E\{S_i(g_n) \mid \theta_n\} + \nabla E_{\bar{\theta}_n} \cdot (E_{\hat{\theta}_n} - E_{\theta_n}) \ .$$

By re-arranging, we see that

$$
\begin{aligned}
|E\{S_i(g_n) \mid \hat{\theta}_n(\mathbf{y})\} - E\{S_i(g_n) \mid \theta_n\}| &= |\sum_{i=1}^{n} \partial_i E_{\tilde{\theta}_n} (\hat{\theta}_{n,i} - \theta_{n,i})| \\
&\leq \sum_{i=1}^{n} |\partial_i E_{\tilde{\theta}_n} (\hat{\theta}_{n,i} - \theta_{n,i})| \\
&\leq \sup_{\tilde{\theta}_n} \sum_{i=1}^{n} |\partial_i E_{\tilde{\theta}_n}| \cdot |(\hat{\theta}_{n,i} - \theta_{n,i})|
\end{aligned}
$$

Under Condition 3, we have that $\sup_{\theta_n} \partial_i E_{\theta_n} \leq C/n$ for some constant C, so we can then upper bound

$$
|E\{S_i(g_n) \mid \hat{\theta}_n(\mathbf{y})\} - E\{S_i(g_n) \mid \theta_n\}| \leq \frac{C}{n} \sum_j |\hat{\theta}_i(n) - \theta^\star(n)_i| \ ,
$$

and this last term is $o_P(1)$ by Condition 1 of the theorem. This completes the proof. □

S4. Proofs of Taxonomy Results (Corollaries 1 and 2)

Proof of Corollary 1. This is straightforward to calculate:

$$
E\left[\left\{E(g_{ij}) - g_{ij}^*\right\}^2\right] = E\{E(g_{ij})^2 - 2E(g_{ij})g_{ij}^* + g_{ij^*}^2\} = p_{ij}^2(\theta) - 2p_{ij}(\theta)g_{ij}^* + \left(g_{ij}^*\right)^2
$$

which completes the proof. □

Proof of Corollary 2. To prove Corollary 2, we need to verify the three conditions from Theorem 4 in the main paper. We first verify condition 1. This condition requires that the average error $1/n \sum_{i=1}^{n} |\hat{\theta}_i - \theta_i^\star| = o_P(1)$. This is true for the estimators from Theorems 1, 2, and 3 since we have shown that the maximum error converges to zero in probability.

We now turn to proving that Condition 3 of 4 is satisfied. That is, we want to verify that $|E\{S_i(g_n) \mid \theta_n^\star\} - S_i(g_n^\star)| \xrightarrow{p} 0$ as $n \to \infty$.

For part 1, density, we have

$$
\begin{aligned}
\sum_{j \in \{1,\ldots,n\}, j \neq i} \frac{\mathrm{var}(g_{ij})}{(n-1)^2} &= \sum_{j \in \{1,\ldots,n\}, j \neq i} \frac{p_{ij}(\theta)\left(1 - p_{ij}(\theta)\right)}{(n-1)^2} \\
&\leq \sum_{j \in \{1,\ldots,n\}, j \neq i} \frac{1}{(n-1)^2} = \frac{1}{n-1} \to 0
\end{aligned}
$$

so the Kolmogorov condition is satisfied and

$$
\mathbb{P}\left\{\lim_{n \to \infty} \frac{d_i}{n} = \frac{E\left(d_i\right)}{n}\right\} = 1
$$

which satisfies the conditions of Theorem 4.

In part 2 we turn to diffusion centrality. Recall that.

$$
DC_i\left(g; q_n, K\right) = \sum_j \left\{\sum_{t=1}^{K} \left(q_n g\right)^t\right\}_{ij} = \sum_j \sum_{t=1}^{K} \frac{C^t}{n^t} \sum_{j_1,\ldots,j_{t-1}} g_{ij_1} \cdots g_{j_{t-1}j}.
$$

For any t, we have

$$
\begin{aligned}
\mathrm{var}\left(\frac{1}{n^t} \sum_j \sum_{j_1,\ldots,j_{t-1}} g_{ij_1} \cdots g_{j_{t-1}j}\right) &= \frac{1}{n^{2t}} \sum_j \sum_{j_1,\ldots,j_{t-1}} \mathrm{var}(g_{ij_1} \cdots g_{j_{t-1}j}) \\
&+ \frac{1}{n^{2t}} \sum_j \sum_{j_1,\ldots,j_{t-1}} \sum_k \sum_{k_1,\ldots,k_{t-1}} \mathrm{cov}(g_{ij_1} \cdots g_{j_{t-1}j}, g_{ik_1} \cdots g_{k_{t-1}k})
\end{aligned}
$$

where $j_0 = k_0 = i$ and $j_s = j, k_s = k$. $\mathrm{var}(g_{ij_1} \cdots g_{j_{t-1}j})$ has variance $\prod_{s=1}^{t} p_{j_{s-1}j_s}\left(1 - \prod_{s=1}^{t} p_{j_{s-1}j_s}\right) \leq 1$ and $\mathrm{cov}(g_{ij_1} \cdots g_{j_{t-1}j}, g_{ik_1} \cdots g_{k_{t-1}}$ 1. In order for $\mathrm{cov}(g_{ij_1} \cdots g_{j_{t-1}j}, g_{ik_1} \cdots g_{k_{t-1}k}) \neq 0$, $g_{ij_1} \cdots g_{j_{t-1}j}$ and $g_{ik_1} \cdots g_{k_{t-1}k}$ need to have at least one edge in common.

Notice that $g_{ij_1}\cdots g_{j_{t-1}j}$ has n^t combinations since i is given. Therefore, given a fixed common edge that $g_{ij_1}\cdots g_{j_{t-1}j}$ and $g_{ik_1}\cdots g_{k_{t-1}k}$ share, $g_{ij_1}\cdots g_{j_{t-1}j}$ has n^{t-2} free choices of actors in the path, and $g_{ik_1}\cdots g_{k_{t-1}k}$ also has n^{t-2} free choices of actors in the path. Therefore, for a given fixed common edge, there are $n^{2(t-2)}$ non-zero covariance terms. Since there are n^2 choices of a common edge, there are a total of n^{2t-2} non-zero covariance terms. Therefore,

$$\mathrm{var}\left(\frac{1}{n^t}\sum_j \sum_{j_1,\ldots,j_{t-1}} g_{ij_1}\cdots g_{j_{t-1}j}\right) \leq \frac{n^t + n^{2t-2}}{n^{2t}}.$$

Let $DC_{i,t} = \frac{1}{n^t}\sum_j \sum_{j_1,\ldots,j_{t-1}} g_{ij_1}\cdots g_{j_{t-1}j}$, we have

$$\mathbb{P}\left\{DC_{i,t} - E(DC_{i,t}) \geq \epsilon\right\} \leq \frac{n^t + n^{2t-2}}{n^{2t}\epsilon^2} \text{ by Chebyshev's inequality}$$

$$\mathbb{P}\left\{DC_{i,t} - E(DC_{i,t}) \mid < \epsilon\right\} \geq 1 - \frac{n^t + n^{2t-2}}{n^{2t}\epsilon^2} \to 1 \text{ as } n \to \infty$$

Therefore, $DC_{i,t}$ goes in probability to $\mathrm{E}(DC_{i,t})$ as $n \to \infty$ and, by continuous mapping theorem,

$$DC_i\left(g; q_n, K\right) = \sum_{t=1}^{K} C^t \times DC_{i,t}$$

tends to $E(DC_i\left(g; q_n, K\right))$ in probability.

For part 3, clustering, the argument is identical to the convergence of clustering in Erdos-Renyi graphs because every link is conditionally edge independent. Let $N(i)$ denote the set of neighbors of actor i and $N(i)$ denote the size of neighbors, then

$$clustering_i(g) = \frac{\sum_{j,k\in N(i)} g_{jk}}{N(i)\cdot\{N(i) \mid -1\}}$$

Similar to the proof for density, we have

$$\sum_{j,k\in N(i)} \frac{\mathrm{var}(g_{jk})}{[\mid N(i) \mid \times\{\mid N(i) \mid -1\}]^2} = \sum_{j,k\in N(i)} \frac{p_{jk}(\theta)\left(1-p_{jk}(\theta)\right)}{[\mid N(i) \mid \times\{\mid N(i) \mid -1\}]^2}$$
$$\leq \sum_{j,k\in N(i)} \frac{1}{[\mid N(i) \mid \times\{\mid N(i) \mid -1\}]^2} = \frac{1}{\mid N(i) \mid \times\{\mid N(i) \mid -1\}} \to 0$$

so the Kolmogorov condition is satisfied and $clustering_i(g)$ goes in probability to

$$E_{z_j,\nu_j,z_k,\nu_k|j,k\in N(i)}\{\mathbb{P}(g_{jk} = 1 \mid \nu_j, \nu_k, z_j, z_k)\}$$

as n tends to infinity.

Finally, we now verify Condition 2 of Theorem 4 of the main paper.

The degree of a node i is $S_i(g_n) = 1/(n-1)\sum_{j\neq i} p_{ij}(\theta)$. In this case, for any k,

$$\partial_k E\{S_i(g_n) \mid \theta_n\} = \frac{1}{n-1}\frac{d}{d\theta_k}p_{ik}(\theta)$$

So, supposing that $\frac{d}{d\theta_k}p_{ik}(\theta)$ is uniformly bounded, which we assume in the statement of Corollary 2, we can conclude for some constant C that $\frac{d}{d\theta_k}p_{ik}(\theta) \leq C$ uniformly over k. We can then conclude that $\partial_k E\{S_i(g_n) \mid \theta_n\} \leq C/(n-1)$ for some constant C, so Condition 2 holds for the degree statistic. A similar argument applies to the clustering coefficient of a node, defined as

$$S_i(g_n) = \frac{1}{\binom{N_i}{2}} \sum_{j,k\in N_i} g_{ij}g_{jk}$$

where N_i is the set of neighbors of node i: $N_i = \{j : g_{ij} = 1\}$.

We finally look at the centrality parameter of a node. We only look at the case of $T = 2$, since the argument for $T > 2$ is similar. We begin by computing $E\{S_i(g_n) \mid \theta_n\}$, which is equal to

$$E\{S_i(g_n) \mid \theta_n\} = \sum_j \frac{C}{n}E[A_{ij}] + \sum_j \frac{C^2}{n^2}E\{[A^2]_{ij}\}\ .$$

where A^2 is the matrix square of the matrix A and A is the adjacency matrix of the graph g. We are interested in the derivative of $E\{S_i(g_n) \mid \theta_n\}$. Supposing that $\frac{d}{d\theta_k} p_{ik}(\theta)$ is uniformly bounded, the derivative of the first term satisfies Condition 3. So we now turn to the second sum and expand

$$E\{A_{ij}\}^2 = E\left\{\sum_k A_{ik}A_{kj}\right\} = \sum_k E\{A_{ik}A_{kj}\} = \sum_k E\{A_{ik}\}E\{A_{kj}\} = \sum_k p_{ik}(\theta)p_{kj}(\theta) \ .$$

Under the same assumption that the derivative $\frac{d}{d\theta_k} p_{ik}(\theta)$ is uniformly bounded, we can conclude that the second sum is also satisfies Condition 2.

Thus, we have shown that the three statistics in Corollary 2 satisfy the three conditions in 4, which completes the proof. □

S5. Proof of Consistency of OLS estimators in many networks setting (Theorem 5)

Proof of Theorem 5. We consider the case where there is no intercept ($\alpha = 0$) to simplify the calculations, but the same argument applies to the case where $\alpha \neq 0$.

We begin by expanding

$$O_r = \beta E\{S_r(g_n) \mid \hat{\theta}_r(n)\} + \epsilon_r = \beta S_r^\star + \big(\epsilon_r + \beta E\{S_r \mid \hat{\theta}_r(n)\} - \beta E\{S_r \mid \theta_r\} + \beta E\{S_r \mid \theta_r\} - \beta S_r^\star\big)$$

Let $\tilde{\epsilon}_{n,r} = \big(\epsilon_r + \beta E\{S_r \mid \hat{\theta}_r(n)\} - \beta E\{S_r \mid \theta_r\} + \beta E\{S_r \mid \theta_r\} - \beta S_r^\star\big)$. Now, by using the analytic expression for the OLS estimator, we have that

$$\begin{aligned}
|\hat{\beta} - \beta| &= \frac{1}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R |E\{S_r \mid \hat{\theta}_r(n)\}\tilde{\epsilon}_{n,r}| \\
&= \frac{1}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}\big(|\epsilon_r + \beta E\{S_r \mid \hat{\theta}_r(n)\} - \beta E\{S_r \mid \theta_r\} + \beta E\{S_r \mid \theta_r\} - \beta S_r^\star|\big) \\
&\leq \frac{1}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)^2} \sum_{r=1}^R |E\{S_r \mid \hat{\theta}_r(n)\}\epsilon_r| + \frac{\beta}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R |E\{S_r \mid \hat{\theta}_r(n)\}\left(E\{S_r \mid \hat{\theta}_r(n)\} - E\{S_r \mid \theta_r\}\right)| + \\
&\frac{\beta}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R |E\{S_r \mid \hat{\theta}_r(n)\}\left(E\{S_r \mid \theta_r\} - S_r^\star\right)| \\
&= I + II + III \ .
\end{aligned}$$

Now, I is $o_P(1)$ assuming that $E(\epsilon_r | E\{S_r \mid \hat{\theta}_r(n)\}) = 0$. Now, let us look at the second term,

$$II = \frac{1}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\} \times |E\{S_r \mid \hat{\theta}_r(n)\} - E\{S_r \mid \theta_r\}| \ ,$$

and the third term is

$$III = \frac{1}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\} \times |E\{S_r \mid \theta_r\} - S_r^\star|$$

For the third term, supposing that $E\{S_r \mid \hat{\theta}_r(n)\} \leq C$, I can upper bound

$$III \leq \frac{C}{R^{-1}\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \frac{1}{R} \sum_{r=1}^R |E\{S_r \mid \theta_r\} - S_r^\star|$$

Now suppose that that $E\{S_r^\star \mid \theta\}$ has finite mean. We then can then conclude that

$$III \leq \frac{C}{R^{-1}\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \frac{1}{R} \sum_{r=1}^R |E\{S_r \mid \theta_r\} - S_r^\star| \ .$$

By Hoeffding's inequality, we can conclude that the average $\frac{1}{R}\sum_{r=1}^R |E\{S_r \mid \theta_r\} - S_r^\star| = o_P(1)$, and so by Slutksy's lemma, we can conclude that $III = o_P(1)$ as $n, R \to \infty$.

We now move to the second term II. Using a Taylor series expansion, we can write

$$\begin{aligned}
E\{S_r \mid \hat{\theta}_r(n)\} - E\{S_r \mid \theta_r(n)\} &= D^T(\bar{\theta}_n)||\hat{\theta}_r(n) - \theta_r(n)|| \\
&= \sum_{i=1}^n \partial_i E\{S_r \mid \bar{\theta}_r(n)\}|\hat{\theta}_r(n) - \theta_r(n)|_i
\end{aligned}$$

for some sequence of intermediate values $\bar{\theta}_n$. So,

$$
\begin{aligned}
II &\leq \frac{1}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\} \sum_{i=1}^n \partial_i E\{S_r \mid \bar{\theta}_r(n)\} \times |\hat{\theta}_r(n) - \theta_r(n)|_i \\
&\leq \frac{C}{\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \sum_{r=1}^R \frac{1}{n} \sum_{i=1}^n |\hat{\theta}_r(n) - \theta_r(n)|_i \\
&= \frac{C}{R^{-1}\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \frac{1}{R} \sum_{r=1}^R \sum_{i=1}^n |\hat{\theta}_r(n) - \theta_r(n)|_i
\end{aligned}
$$

where the first inequality follows from the Taylor series expansion and the second inequality follows from the assumptions of this theorem. Supposing that that $E(E\{S_r \mid \hat{\theta}_r(n)^2) < \infty$, we bound

$$
II \leq \frac{C}{R^{-1}\sum_{r=1}^R E\{S_r \mid \hat{\theta}_r(n)\}^2} \max_{1 \leq r \leq R} \sum_{i=1}^n |\hat{\theta}_r(n) - \theta_r(n)|_i
$$

Under the assumptions of the theorem, we have that $\max_{1\leq r \leq R} \sum_{i=1}^n |\hat{\theta}_r(n) - \theta_r(n)|_i = o_P(1)$, so we conclude that $|\hat{\beta}_{n,R} - \beta| = o_P(1)$, as claimed.

To prove that the estimator $\hat{\gamma}_{n,R}$ is consistent, the argument is nearly identical. To see why, we simple re-arrange the supposed data generating model:

$$
S_r^\star = \alpha + \gamma T_r + \epsilon_r - E\{S_i(g_n) \mid \hat{\theta}_r(n)\} + S_r^\star \ . \qquad \text{[S.11]}
$$

The same argument applies to show that the OLS estimates of γ are also consistent under the conditions of the theorem. ☐

S6. Checking conditions of Theorem 5 for common network statistics (Theorem 6)

Proof of Theorem 6. We only prove the case for the density. The arguments for the other two statistics are similar.

From the proof of Theorems 1, 2, 3, we showed that for any network, each estimator $\hat{\theta}_{i,r}(n)$ satisfies an exponential concentration inequality, and by taking a union bound over all nodes in a network, we see that

$$
\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n |\hat{\theta}_{i,r}(n) - \theta_{i,r}^\star| > \epsilon\Big) \leq \mathbb{P}\Big(\max_{1\leq i \leq n} |\hat{\theta}_{i,r}(n) - \theta_{i,r}^\star| > \epsilon\Big) \leq nC\exp(-C'\epsilon^2 n) \ .
$$

for some constants C and C'. By taking a union bound over all R villages, we conclude that

$$
\mathbb{P}\Big(\max_{1\leq r\leq R} \frac{1}{n}\sum_{i=1}^n |\hat{\theta}_{i,r}(n) - \theta_{i,r}^\star| > \epsilon\Big) \leq Rn\exp(-\epsilon^2 n) \ .
$$

Under the assumptions of the theorem, we have that $Rn\exp(-n) \to 0$, so Condition 2 holds. We now discuss Condition 3 of Theorem 5. One way to satisfy this is to require that the network statistic for each network is the same (i.e., we are considering just the centrality of a set of nodes). In this case, since the network statistic $S_{i,r}$ satisfies the required derivative condition, per Theorem 4, we can then conclude that the maximum also satisfies such a derivative condition. This completes the proof. ☐

S7. Results using fully-elicited graphs

In this section we present additional results using fully-elicited, observed graphs. We use data from (9), which consists of completely observed graphs from 75 villages in rural India. The goal of these results is two-fold. First, we aim to demonstrate that our results hold in networks that have the level of sparsity and complexity that a user could reasonably find in practice. We explore the notion of sparsity further in Section S11 Second, we aim to show that the performance of our method improves as the graph size increases, as indicated by our results.

In each village, about one-third of respondents were asked ARD questions. (7) compare statistics estimated with ARD from these graphs with the same statistics calculated using the complete graph. We leverage these results and present a different aspect, how the MSE changes as the size of the graph grows. We present results for individual-level statistics from these graphs and compute MSE across individuals. Figure S2 presents these results. Each point in the figure represents one village. The size of the village is along the x-axis. In practice we report the number of nodes that have ARD as the axis labels, which is about 1/3 of the total number of nodes in the graph. For nodes that do not have ARD responses we use the procedure described in (7).

S8. Additional simulation results with estimated formation model parameters

In this section we present additional simulation results to complement the simulations we present in the main text. We present results when the parameters are estimated using the procedure in (7), rather than assumed to be consistently estimated. These simulations are presented in Figures S3, S4, and S5. The results we present here use the same simulation setup as Figure 2 in the main text.

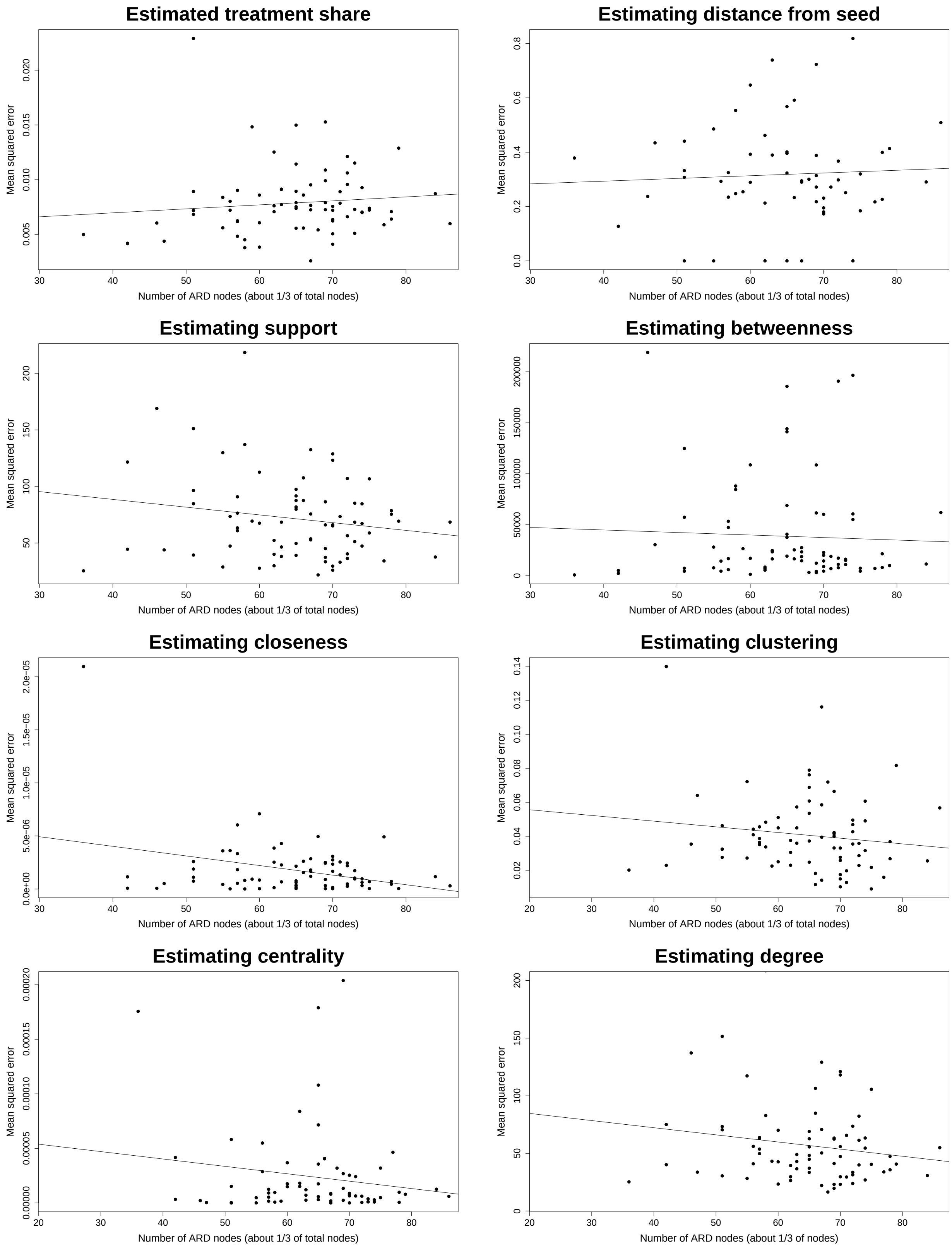


Fig. S2. MSE and graph size. Each plot shows the MSE (computed across nodes) plotted as a function of the number of respondents who received ARD using data from (9).

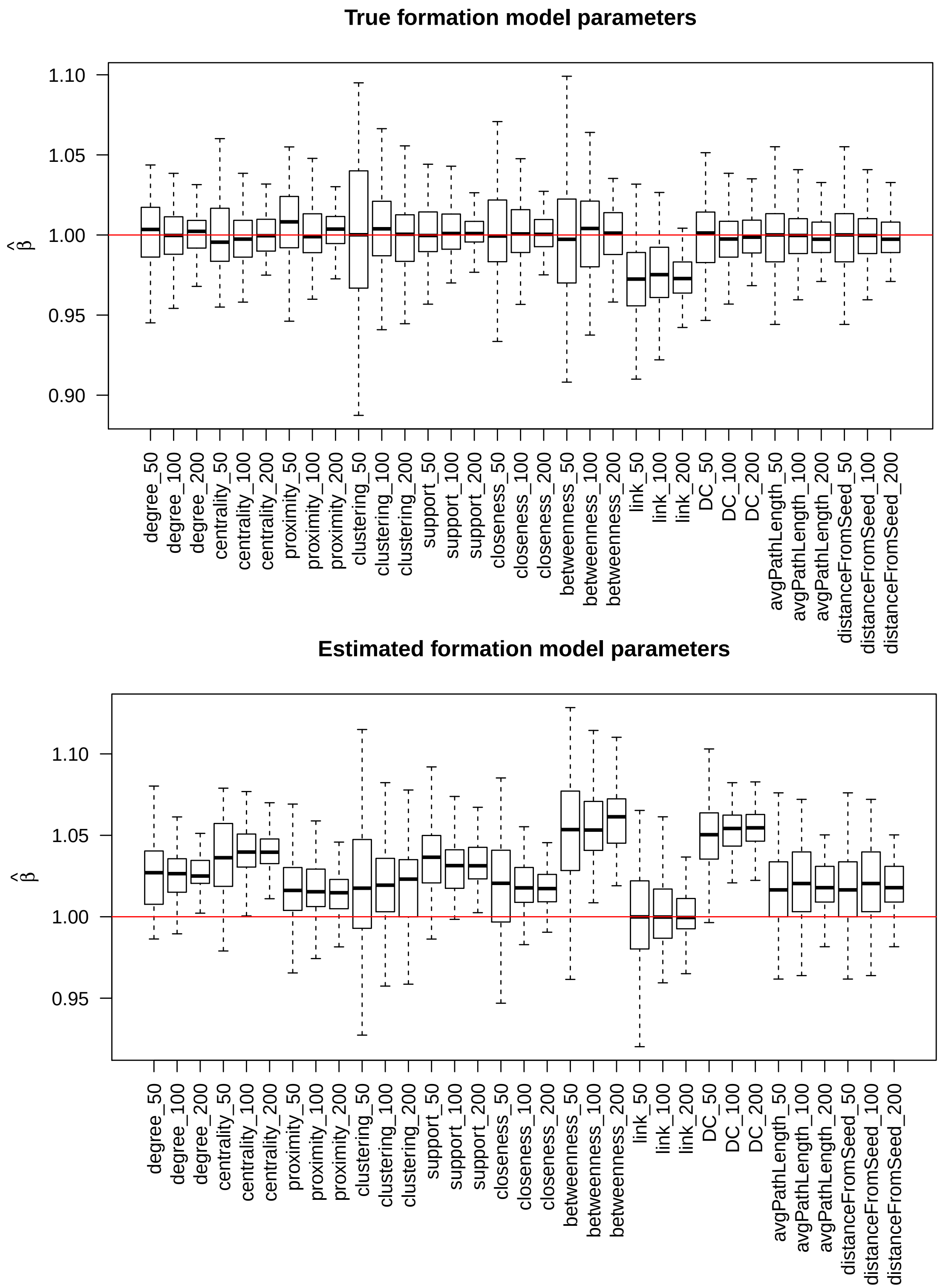
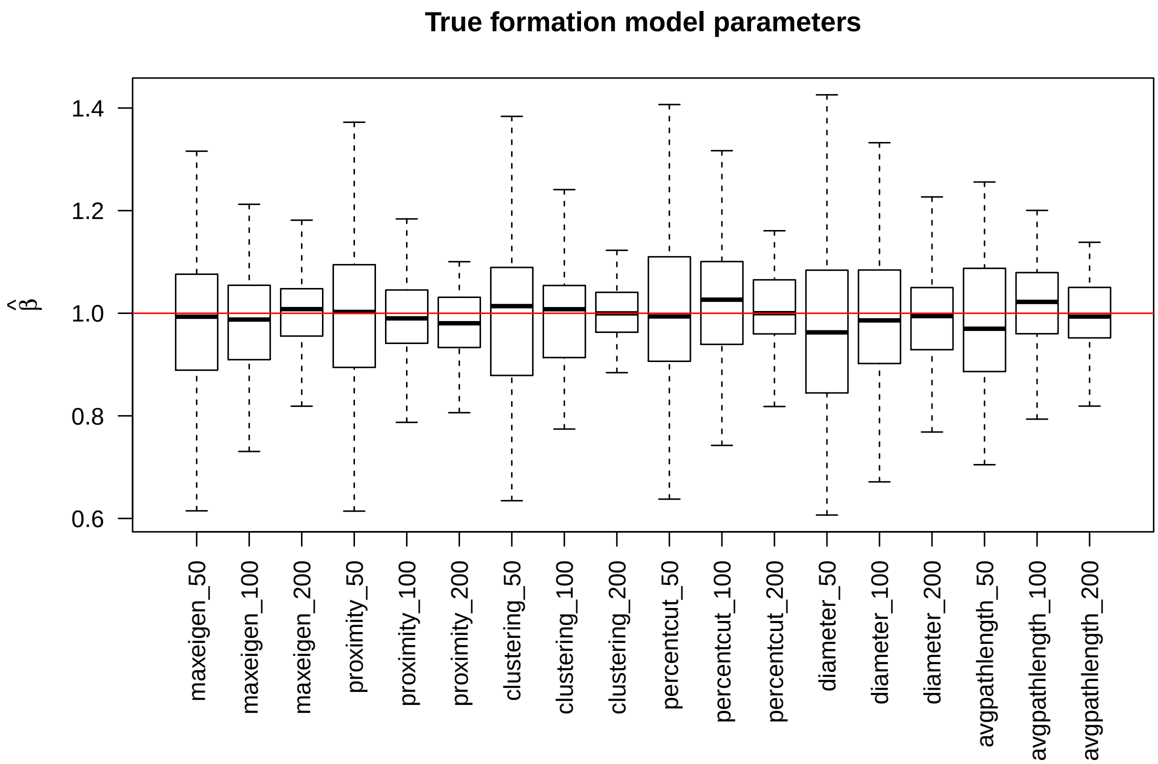


Fig. S3. Boxplot of $\hat{\beta}$ for β in regression $y_{ij,r} = \alpha + \beta \bar{S}_{ij,r} + \epsilon_r$, where $S_{ij,r}$ and $\bar{S}_{ij,r}$ represent a true and mean individual-level measure, respectively. Each box represents the distribution of $\hat{\beta}$ for one measure and use of R=50, 100 or 200 networks in regression. 50 actors and 1000 pairs (for link) are randomly selected for each network. The middle line of the boxplot denotes median, and borders of the boxes denote first and third quartile. The red line denotes the true $\beta = 1$ used to generate $y_{ij,r} = \alpha + \beta S^*_{ij,r} + \epsilon_r$ in the simulation. These results corroborate the theoretical intuition developed in Theorems 4 and 5.



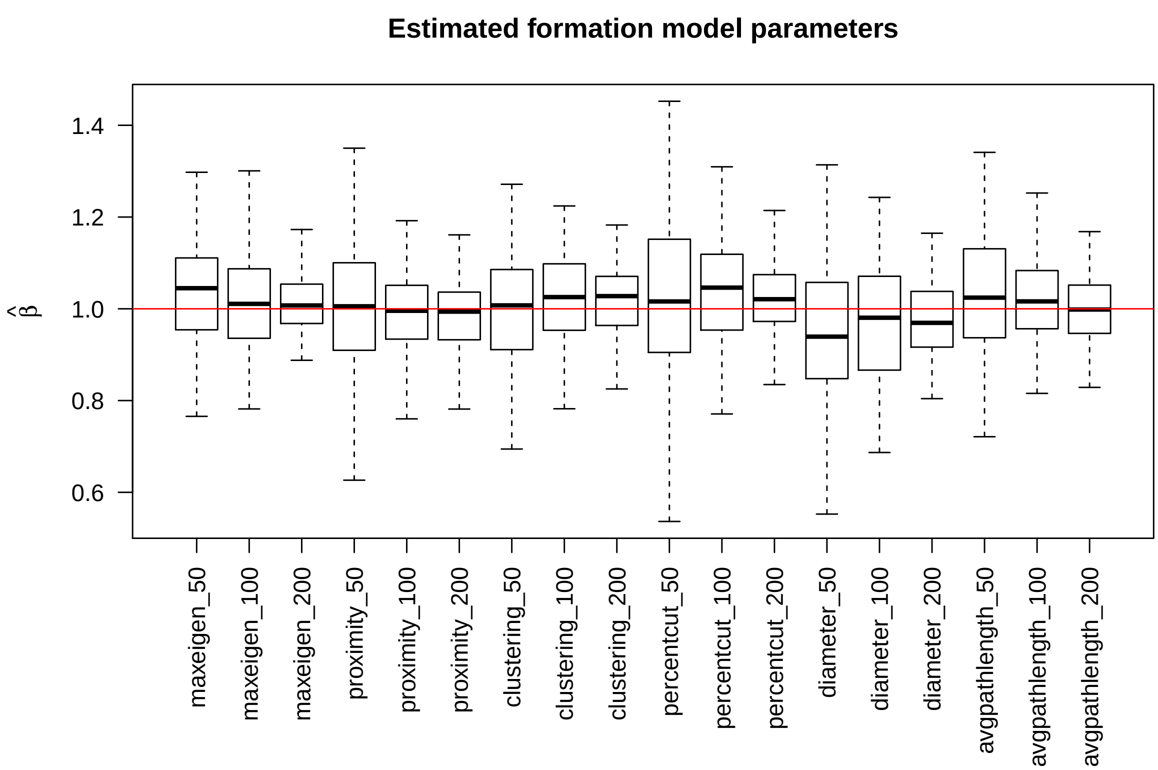


Fig. S4. Boxplot of $\hat{\beta}$ for β in regression $y_r = \alpha + \beta\bar{S}_r + \epsilon_r$, where S_r and $\bar{S}_r$ represent a true and mean network-level measure, respectively. Each box represents the distribution of $\hat{\beta}$ for one measure and use of R=50, 100 or 200 networks in regression. The middle line of the boxplot denotes median, and borders of the boxes denote first and third quartile. The red line denotes the true $\beta = 1$ used to generate $y_r = \alpha + \beta S_r^* + \epsilon_r$ in the simulation. These results corroborate the theoretical intuition developed in Theorems 4 and 5.

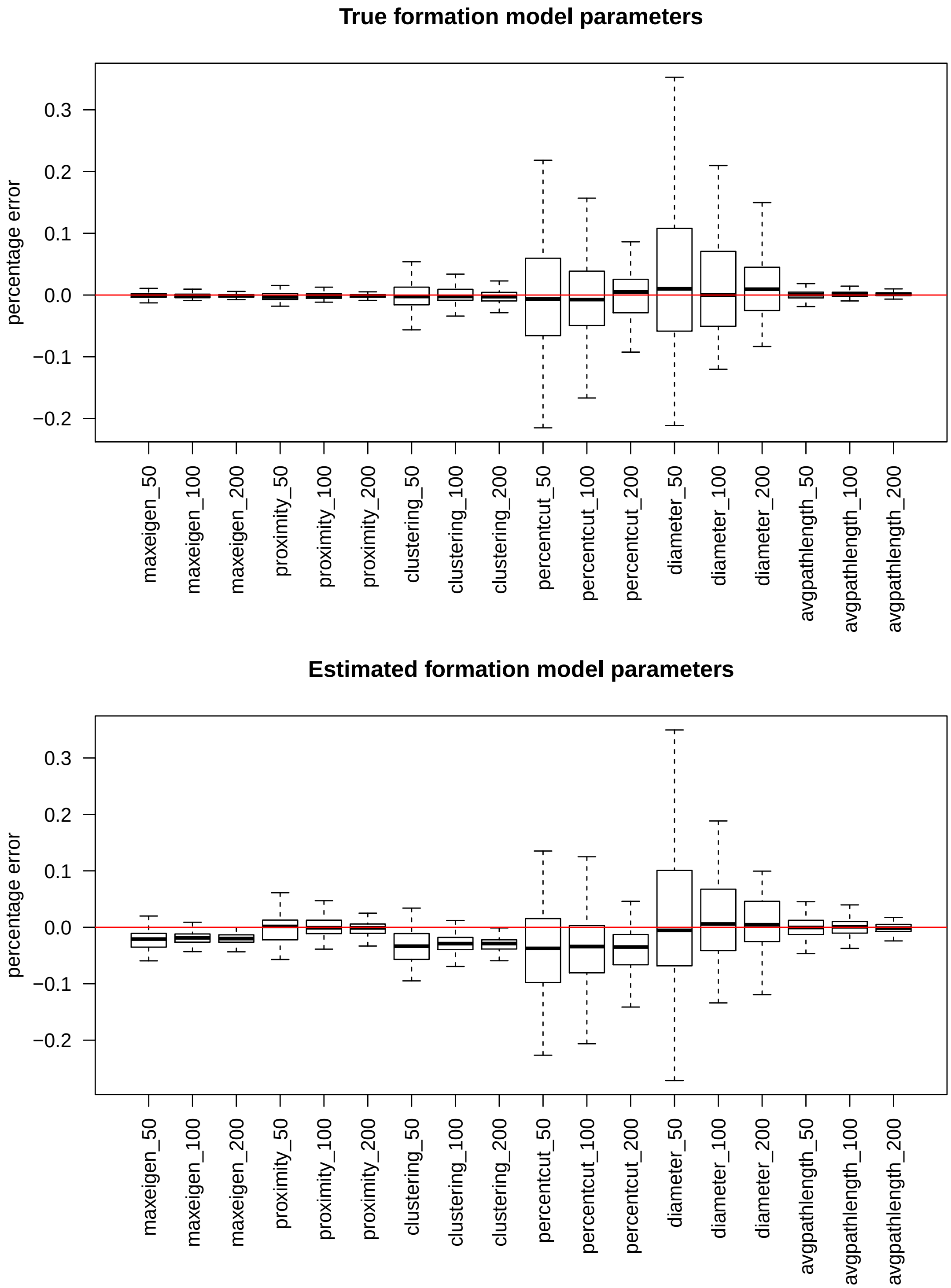


Fig. S5. Boxplot of percentage errors of $\hat{\gamma}$ for γ in regression $\bar{S}_r = \alpha + \gamma T_r + \epsilon_r$, where S_r and $\bar{S}_r$ represent a true and mean network-level measure, respectively. Each box represents the distribution of percentage errors for one measure and use of R=50, 100 or 200 networks in regression. The middle line of the boxplot denotes median, and borders of the boxes denote first and third quartile. These results corroborate the theoretical intuition developed in Theorems 4 and 5.

S9. Simulations to Demonstrate Consistency of Latent Space Model Parameter Estimators

In this section, we study simulation experiments to we show that the estimates of $z_i^\star$ and $\nu_i^\star$ are consistent as $n \to \infty$.

We start with the estimates of the node locations. To do this, we create two group centers $\mu_1 = (2, 2)$ and $\mu_2 = (-2, -2)$ and set $z_0 = (0, 0)$. Our goal is to estimate the location of z_0. In Figure S6, we plot a sample realization of the z_i and z_0 for $n = 500$.

We assign n nodes to be in group 1, and n nodes to be in group 2. Given these group memberships c_i, we draw

$$z_i \mid \{c_i = j\} \sim N\left(\mu_j, \frac{1}{3} I_2\right) \quad j = 1, 2\ .$$

where I_2 is the 2×2 identity matrix. We then create generate edges between the node at location z_i and z_0 by defining

$$P_i = \exp(-||z_i - z_0||) = \exp(-||z_i||)\ .$$

where the second equality follows since $z_0 = (0, 0)$. We then generate the edges between nodes in groups 1 and 2 and the node at z_0 in this way:

$$\begin{aligned} G_{i1} &= \text{Bernoulli}(P_i), \quad c_i = 1 \\ G_{i2} &= \text{Bernoulli}(P_i), \quad c_i = 2\ . \end{aligned}$$

The ARD responses are then $y_{i1} = \sum_{i=1}^{n} G_{i1}$ and $y_{i2} = \sum_{i=n+1}^{2n} G_{i2}$. We then estimate the node location z_0 by the estimation procedure described above. In particular, the estimate $\hat{z}_i$ solves $\hat{z}_i = G_1(a)$, where $a = \log(Y_{i1}/n) - \log(Y_{i2}/n)$. We repeat the above process 25 times for each value of $n = 50, 100, 500, 1000, 10^4$. In Figure S7, we plot $||\hat{z}_i - z_i|| = ||\hat{z}_i||$. We see that the norm is decreasing as n increases.

To demonstrate the consistency claim for the node effect estimate $\hat{\nu}_i$, we simulate n locations $z_i \sim N\left((2, 2), \frac{1}{3} I_2\right)$ and $\nu_i^\star \overset{\text{i.i.d.}}{\sim} \text{Unif}(-2, 0)$. We then let $\nu_i^\star = -1$. Our estimate of the node effects is, recalling (S.3), the $\hat{\nu}_i$ that solves

$$\frac{y_{ik}}{n_k} = E\{\exp(\nu^\star)\} \exp(\hat{\nu}_i) E[\exp\{-d(z_i, z)\}]\ ,$$

where $z \sim F(\mu_k^\star, \sigma_k^\star)$. We suppose that the terms z_i, $E\{\exp(\nu^\star)\}$ and $\mu_k^\star, \sigma_k^\star$ are known, which allows us to solve for the estimate $\hat{\nu}_i$. We repeat this process 100 times for $n = 250, 500, 1000, 10^4$. In Figure S8, we plot the estimation error and see that as n increases, the error decreases.

S10. Supplemental results used to prove Theorem 3

In this section, we prove Lemma S2.1 which is used to prove Theorem 3. To do that, we introduce the pseudo-log likelihood of the ARD. We note here that maximizing the pseudo-likelihood is equivalent to the method-of-moments (or equivalently, Z-estimator) approach taken in Section S2 but by maximizing the pseudo log likelihood, we are able to use the classical M-estimator results to conclude consistency (5).

We now discuss the pseudo-likelihood of the ARD. As described above, the data we observe, when conditioned on the ego's parameters and marginalizing over the alters' parameters, are simply Binomial draws. We can write the log-likelihood for the number of links that i has to a random set of n_k members of group k as

$$\log f(y_{ik} \mid \nu_i, z_i, \eta) = \log\left\{\binom{n_k}{y_{ik}}\right\} + y_{ik}\log(p_{ik}) + (n_k - y_{ik})\log(1 - p_{ik}). \qquad \text{[S.12]}$$

for an arbitrary $\nu_i, z_i,\ \eta$.

We can build our target objective function by summing up over all k traits for each node and then all nodes

$$\sum_{i=1}^{m}\sum_{k=1}^{K} \log f(y_{ik} \mid \nu_i, z_i, \eta).$$

For each i, the counts of links across groups are independent conditional on the latent positions. We describe this as the pseudo-likelihood because the full likelihood also accounts for correlation between $Y_{ik(j)}$ and $Y_{jk(i)}$, where $k(i)$ is person i's group. Nonetheless, this pseudo-likelihood delivers consistent estimates, similar to other recent work in consistent estimators for graph models. See (10) and its references for a discussion on this point. In practice, we do not know the parameter $\eta^\star$, which contains the means and variances of the distribution of node locations as well as the expxected value of $\exp(\nu_i)$. Suppose that we have a consistent estimator $\hat{\eta} \xrightarrow{p} \eta^\star$. We can then use this plug-in estimator in place of $\eta^\star$, which leads to the final ARD pseudo-likelihood

$$\hat{\ell}_n(\mathbf{y} \mid \theta) = \sum_{i=1}^{m}\sum_{k=1}^{K} \log f(y_{ik} \mid \nu_i, z_i, \hat{\eta}). \qquad \text{[S.13]}$$

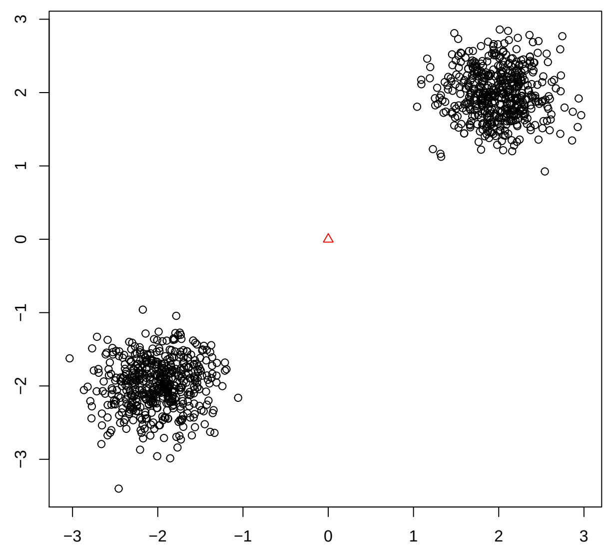


Fig. S6. Plot of $n = 500$ locations (black circle) centered at $(2, 2)$ and $(-2, 2)$. The point at $(0, 0)$ (the red triangle) is the location we want to estimate with the ARD.

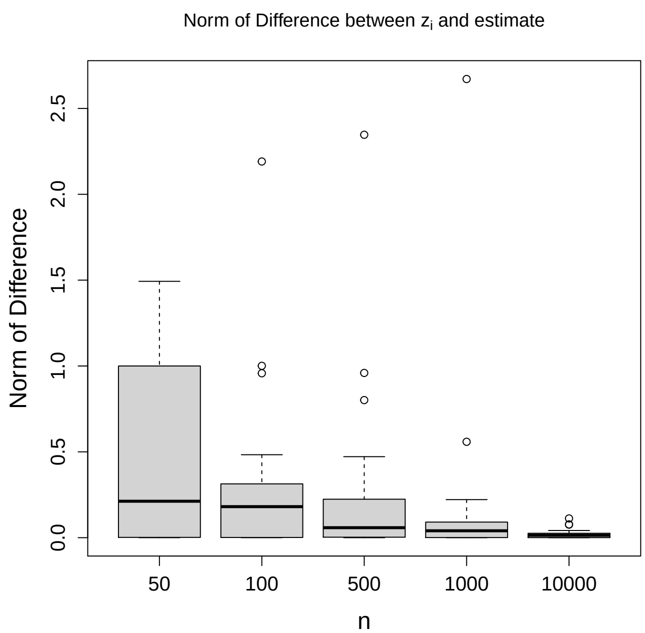


Fig. S7. Norm of difference $\hat{z}_i - z_0$ for various values of n on the x-axis.

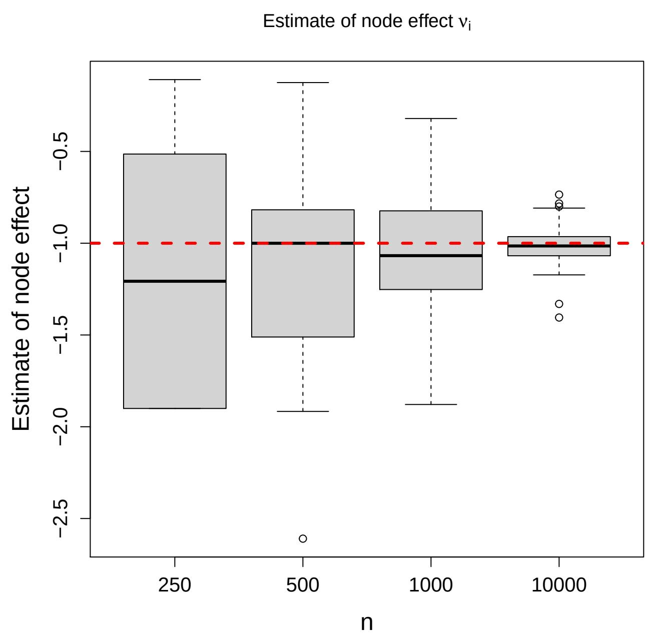


Fig. S8. Estimate of the node effect $\nu_i^\star$ using the esimate defined in (S.3). We set $\nu_i^\star = -1$ and generate estimates of this parameter using various values of n on the other x-axis. As n increases, we see convergence of the estimate to $\nu_i^\star$.

We then define the estimates of the node locations and effects as the maximisers of the following pseudo-likelihood:

$$(\hat{\nu}_1, \ldots, \hat{\nu}_m, \hat{z}_1, \ldots, \hat{z}_m) = \underset{\nu_{[1:m]}, z_{[1:m]}}{\arg\max}\ \hat{\ell}_n \left(\mathbf{y} \mid \nu_{[1:m]}, z_{[1:m]}, \hat{\eta}\right) . \qquad \text{[S.14]}$$

We begin by including the following result, Theorem 5.7 of (5), that allows us to conclude consistency of an M-estimator. This result requires two conditions, which we now state below.

Condition **S10.1.** *For all $\epsilon > 0$,*

$$\sup_{\theta : d(\theta^\star, \theta) \geq \epsilon} Q(\theta) < Q(\theta^\star) .$$

When Θ is compact, which we assume is true in Condition S10.3 below, a sufficient condition for Condition S10.1 to hold is that Q has a unique maximum at $\theta^\star$.

Condition **S10.2** (Uniform law of Large Numbers). *We require that*

$$\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q(\theta)| \xrightarrow{p} 0 .$$

Under these two conditions, we can conclude that any M-estimator of the form $\hat{\theta}_n = \arg\max Q_n(\theta)$ is consistent, in the sense specified below.

Lemma **S10.1** (Theorem 5.7 of (5)). *Let $\hat{Q}_n$ be a sequence of random functions indexed by $\theta \in \Theta$, where (Θ, d) is a metric space. Suppose that Conditions S10.1 and S10.2 hold. Then, $d(\hat{\theta}, \theta^\star) \xrightarrow{p} 0$ as $n \to \infty$.*

There are many ways to verify the uniform law of large numbers result in Condition S10.2. See, among others, (8, 11, 12). In this work, we follow the approach outlined by (8), which requires a compact parameter space, that the functions $\hat{Q}_n$ converge pointwise to $E(\hat{Q}_n)$, and that the functions $\hat{Q}_n$ satisfy a Lipschitz-type condition.

The following two conditions are used in the uniform law of large numbers results from (8).

Condition **S10.3** (Compact Parameter Space). *We suppose that (Θ, d) is a compact metric space.*

Condition **S10.4** (Pointwise Convergence). *For each $\theta \in \Theta$, $\hat{Q}_n(\theta) = \bar{Q}(\theta) + o_P(1)$*

Lemma **S10.2** (Corollary 2.1 of (8)). *Suppose Conditions S10.3 and S10.4 hold and that $\bar{Q}_n$ is equicontinuous. Also suppose that Θ is a metric space with metric $d(\theta, \theta')$ and there exists B_n such that for all $\theta, \theta' \in \Theta$, $|\hat{Q}_n(\theta) - \hat{Q}_n(\theta')| \leq B_n d(\theta, \theta')$ and $B_n = O_P(1)$. Then $\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - \bar{Q}_n(\theta)| = o_P(1)$.*

As (8) points out immediately after Corollary 2.1, if $\bar{Q}_n = E(\hat{Q}_n)$ and $E(B_n)$ is bounded, then we can drop the assumption that $\hat{Q}_n$ is equicontinuous and instead include it as a conclusion to the lemma. In other words, we do not need to check the condition of equicontinuity to use the lemma above.

Lemma **S10.3.** *The likelihood function of the data y_{ik}, conditioned on node i's parameters, which we denote by $f(\nu_i, z_i)$, from the proof of Lemma S2.1 has a unique maximum at $(\nu_i^\star, z_i^\star, \eta^\star)$ for sufficiently large K.*

Proof. By the information decomposition, and again using f to denote the likelihood of y_{ik} given node i's parameters, we have that

$$E[\log\{f(y_{ik} \mid \nu_i, z_i)\}] = H_{ik}(\theta^\star) - KL_{ik}(\theta \mid \theta^\star) .$$

where H is the entropy of $y_{ik} \mid \nu_i^\star, z_i^\star$ and KL is the KL-divergence between $y_{ik}|\nu_i^\star, z_i^\star$ and $y_{ik}|\nu_i, z_i$. See (13) for more information on this decomposition.

So to maximize the $E[\log\{f(y_{ik} \mid \nu_i, z_i)\}]$, we need to minimize the KL divergence. Hence, by summing over $k = 1, \ldots, K$,

$$\sum_{k=1}^{K} KL_k(\theta \mid \theta^\star) = \sum_{k=1}^{K} \log\left\{\frac{p_{ik}(\nu_i, z_i)}{p_{ik}(\nu_i^\star, z_i^\star)}\right\} n_k p_{ik}(\nu_i, z_i) + \log\left\{\frac{1 - p_{ik}(\nu_i, z_i)}{1 - p_{ik}(\nu_i^\star, z_i^\star)}\right\} n_k\{1 - p_{ik}(\nu_i, z_i)\} .$$

Now, note first that the KL divergence is always greater than or equal to zero. Second, the KL divergence is zero if and only if $\theta = \theta^\star$. Note that there are just two parameters ν_i and z_i. For any $k = 1, \ldots, K$, we define the set A_k to be

$$A_k = \{(\nu_i, z_i) : \exp(\nu_i) H_k(z_i) E\{\exp(\nu)\} = p_{ik}(\nu_i^\star, z_i^\star)\} .$$

In words, A_k is the set of parameters (ν_i, z_i) that lead to the same probability $p_{ik}(\nu_i^\star, z_i^\star)$. Since the KL divergence is always greater than or equal to zero, with equality if and only if the parameters are equal, we see that $\bigcap_{k=1}^K A_k$ is the set of maximizers of the function f.

Clearly, $(\nu_i, z_i) \in A_k$ for each k and thus $(\nu_i^\star, z_i^\star) \in \bigcap_{k=1}^K A_k$. To argue that f has a unique maximum at $(\nu_i^\star, z_i^\star)$, we now need to argue that $\{(\nu_i^\star, z_i^\star)\} = \bigcap_{k=1}^K A_k$. Supposing that $p_{ik}(\nu_i^\star, z_i^\star) \neq p_{ik'}(\nu_i^\star, z_i^\star)$ for some $k \neq k'$, meaning we have at least two distinct probabilities, then f has a unique maximum. For K sufficiently large, we will have that $\{(\nu_i^\star, z_i^\star)\} = \bigcap_{k=1}^K A_k$. Thus, f has a unique maximum. ☐

Proof of Lemma S2.1. To show consistency of the estimates based on maximizing the pseudo-likelihood, we first note that each pair ν_i, z_i appears in exactly K of the terms in the expression from (S.13). That is,

$$(\hat{\nu}_i, \hat{z}_i) = \arg\max_{\nu, z} \sum_{k=1}^{K} n_k^{-1} \log f(y_{ik} \mid \nu_i, z_i, \hat{\eta})$$

Thus, we will show that each pair $(\hat{z}_i, \hat{\nu}_i)$ converges to the true value. By recalling that $y_{ik} \mid \nu_i^\star, z_i^\star$ is Binomial, we see that

$$\sum_{k=1}^{K} n_k^{-1} \log f(y_{ik} \mid \nu_i, z_i) = \sum_{k=1}^{K} \left\{ n_k^{-1} \log \binom{n_k}{y_{ik}} + \frac{y_{ik}}{n_k} p_{ik} + \left(1 - \frac{y_{ik}}{n_k}\right) \log(1 - p_{ik}) \right\} .$$

To argue consistency, we will use Theorem 5.7 of (5). To simplify the analysis, first note that the term $n_k^{-1} \log \binom{n_k}{y_{ik}}$ does not depend on the parameters, and also $y_{ik} \mid \nu_i, z_i = \sum_{j \in G_k} g_{ij}$, so the maximum pseudo likelihood estimates $(\hat{\nu}_i, \hat{z}_i)$ also satisfy

$$\begin{aligned}(\hat{\nu}_i, \hat{z}_i) &= \arg\max_{\nu, z} \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j \in G_k} \left\{ g_{ij} \log(\hat{p}_{ik}) + (1 - g_{ij}) \log(1 - \hat{p}_{ik}) \right\} \\ &= \arg\max_{\nu, z} \hat{f}_n(\mathbf{y}, \nu_i, z_i, \hat{\eta}) .\end{aligned}$$

We now define the term $\hat{p}$ in the expression above. Given estimates of the structural parameters $E\{\exp(\nu)\}, \mu_k, \sigma_k^2$, we define

$$\hat{p}_{ik} := \exp(\nu_i) \hat{E}\{\exp(\nu)\} \hat{H}_k(z_i)$$

where $\hat{H}_k(z_i) = E[\exp\{-d(z_i, z)\}]$ is computed using z_j drawn iid from $F(\hat{\mu}_k, \hat{\sigma}_k^2)$ and $\hat{E}\{\exp(\nu)\}$ is the estimate of $E\{\exp(\nu)\}$ defined in the previous section.

Define $f_n(\nu_i, z_i) = E\{\hat{f}_n(\mathbf{y}, \nu_i, z_i, \hat{\eta})\}$ and $f(\nu_i, z_i) = \lim_{n \to \infty} f_n(\nu_i, z_i)$. In the defintion of f_n, the expectation is over the distribution of $\mathbf{y}$ (and note that the distribution of $\hat{\eta}$ is also determined by the distribution of $\mathbf{y}$). To see why, see our discussion where we define particular estimates of $\hat{\eta}$ and note that these estimates depend on $\mathbf{y}$. By Lemma S10.3, f has a unique maximum at $(\nu_i^\star, z_i^\star, \eta^\star)$. Thus, since $V \times M \times E$ is compact, it follows that Condition S10.3 is satisfied. To verify Condition S10.2, we first use the triangle inequality to see that $\sup_{\nu_i, z_i} |\hat{f}_n(\mathbf{y}, \nu_i, z_i, \hat{\eta}) - f(\mathbf{y}, \nu_i, z_i, \hat{\eta})|$ is upper bounded by

$$\sup_{\nu_i, z_i} |\hat{f}_n(\mathbf{y}, \nu_i, z_i, \hat{\eta}) - f_n(\mathbf{y}, \nu_i, z_i, \hat{\eta})| + \sup_{\nu_i, z_i} |f_n(\mathbf{y}, \nu_i, z_i, \hat{\eta}) - f(\mathbf{y}, \nu_i, z_i, \hat{\eta})| .$$

The second term, which is deterministic, converges to zero uniformly over all (ν_i, z_i) by the Weierstrassstrass M-test, which we provide for completeness as Lemma S10.4 and state below:

LEMMA **S10.4** (Weierstrass M-test). *Let $f_n(x) = \sum_{i=1}^{n} f_i(x)$ and $f = \lim_n f_n(x)$. Suppose that there exists M_n such that for each n, $|f_n(x)| \le M_n$ for all x and $\sum_{i=1}^{\infty} M_i < \infty$. Then f_n converges uniformly to f.*

Hence this second term converges uniformly in probability over all (ν_i, z_i). We now look at the first term. To show that this converges uniformly in probability to zero, we will use Corollary 2.1 from (8) which for completeness we provide in Section S10. In particular, if we can show (1) that $\hat{f}_n$ converges pointwise to $E(\hat{f}_n)$ and (2) that $\hat{f}_n$ satisfies the Lipschitz inequality

$$|\hat{f}_n(\mathbf{y}, \nu_i, z_i, \hat{\eta}) - \hat{f}_n(\mathbf{y}, \nu_i', z_i', \hat{\eta})| \le B_n d\{(\nu_i, z_i), (\nu_i', z_i')\} , \qquad \text{[S.15]}$$

where $B_n = O_P(1)$, then Condition 2 holds by Corollary 2.1 of (8).

We first show the pointwise convergence. By assumption, $\hat{p}_{ik} = \exp(\nu_i) \hat{\tau} \hat{H}(z_i)$ is a continuous function of its arguments, and since $\hat{\eta} \xrightarrow{p} \eta^\star$, $\hat{p}_{ik} \xrightarrow{p} p_{ik}$ as $n \to \infty$ by the continuous mapping theorem. Also, conditioned on the ego's parameters, $y_{ik}/n_k \xrightarrow{p} p_{ik}$ (by Chebyshev's inequality, since g_{ij} are independent and bounded), so we conclude the pointwise convergence.

To show (S.15), we upper bound the left hand side by $t_1 + t_2$, where

$$\begin{aligned} t_{1k} &= g_{ij} |\log(\hat{p}_{ik}) - \log(\hat{p}_{ik})| \le g_{ij} |\nu_i - \nu_i' + \log \hat{H}(z_i) - \log \hat{H}(z_i')| \\ t_{2k} &= (1 - g_{ij}) |\log(\hat{p}_{ik}) - \log(\hat{p}_{ik})| \le g_{ij} |\nu_i - \nu_i' + \log \hat{H}(z_i) - \log \hat{H}(z_i')| . \end{aligned}$$

By assumption, $\hat{H}$ is Lipschitz in z and so $|\log\{\hat{H}(z_i)\} - \log\{\hat{H}(z_i')\}| \le C d(z_i, z_i')$ for some constant C, so

$$t_{1k} \le g_{ij} \left\{ |\nu_i - \nu_i'| + C d(z_i, z_i') \right\} \le g_{ij} C' d((\nu_i, z_i), (\nu_i', z_i')) ,$$

and a similar argument holds for t_{2k}. Since the left hand side of (S.15) is upper bounded by $\sum_{k=1}^{K} t_{1k} + t_{2k}$, and since $\sum_{j \in G_k} n_k^{-1} g_{ij}$ is $O_P(1)$, we conclude that (S.15) holds and so we conclude by Corollary 2.1 of (8) that Condition 2 holds. It follows from Theorem 5.7 of (5) that the maximum pseudo-likelihood estimator $(\hat{\nu}_i, \hat{z}_i)$ is consistent. ☐

S11. Additional simulations with lower density

We simulated 250 networks of size 250 using a stochastic blockmodel with K=10 traits and C=5 communities under two conditions: (i) degree/density similar to the Banerjee et al. (9) data and (ii) with an average degree/density similar to the average across Table 1 in Chandrasekhar (14). The goal is to understand how the MSE of estimators derived using the proposed method change when using networks that are overall less dense. Figure S9 shows the MSE (scaled by the $1/E(S_i)^2$ to compare across the statistics, where S_i is the network statistic) using density and network size similar to the Banerjee et al. (9) data. The average degree across the villages is 17.38 and the density is 0.081. We see that overall, similar to the results in Figure 1 of the paper, the scaled MSE is small, with the exception of predicting the presence of a single link (consistent with the results presented in the main body of the paper).

In Figure S10 we have simulated networks with an average degree of 9 and the density of 0.036, which is comparable to the averages presented in Table 1 in Chandrasekhar (14), where the average degree is 8.17 and the density is 0.054. In this table, Chandrasekhar (14) reviews the density of networks observed in several different contexts, making this table a potential benchmark for the level of sparsity a researcher may find in practice. We see that the MSE remains small for the statistics that satisfy our taxonomy results, even when we reduce the density well below the average from the Chandrasekhar (14) table.

(a) Individual level statistics.

(b) Graph level statistics.

Fig. S9. MSE results for a stochastic block model on $n = 250$ nodes using $K = 10$ traits and $C = 5$ communities.

(a) Individual level statistics.

(b) Graph level statistics.

Fig. S10. MSE results for a stochastic block model on $n = 250$ nodes using $K = 10$ traits and $C = 5$ communities.

References

1. Bryan S Graham. An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063, 2017.
2. Sourav Chatterjee, Persi Diaconis, and Allan Sly. Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435, 2011.
3. Peter Bickel, Aiyou Chen, and Elizaveta Levina. The method of moments and degree distributions for network models. *Annals of Statistics*, 39(5):2280–2301, 2011.
4. Shane Lubold, Arun Chandrasekhar, and Tyler McCormick. Identifying latent space geometry of network formation models via analysis of curvature. *arXiv preprint arXiv:2012.10559*, 2020.
5. A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
6. Wassily Hoeffding. On the distribution of the number of successes in independent trials. *Annals of Mathemtical Statistics*, 27:713–721, 1956.

7. Emily Breza, Arun G Chandrasekhar, Tyler H McCormick, and Mengjie Pan. Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 2020.
8. Whitney K. Newey. Uniform convergence in probability and stochastic equicontinuity. *Econometric Research Program Research Memorandum No. 342*, 1989.
9. A. Banerjee, A. Chandrasekhar, E. Duflo, and M.O. Jackson. Diffusion of microfinance. *Science*, 341(6144):1–7, 2013.
10. Arash A. Amini, Aiyou Chen, Peter J. Bickel, and Elizaveta Levina. Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics*, 41(No. 4), 2013.
11. Donald WK Andrews et al. Generic uniform convergence. *Econometric Theory*, 8(2):241–257, 1992.
12. Benedikt M. Potscher and Ingmar Prucha. A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica*, May 1989.
13. Thomas M Cover and Joy A Thomas. *Elements of information theory.* John Wiley & Sons, 2012.
14. Arun Chandrasekhar. Econometrics of network formation. *The Oxford handbook of the economics of networks*, pages 303–357, 2016.