ECONOMETRICS OF SAMPLED NETWORKS

ARUN G. CHANDRASEKHAR[‡] AND RANDALL LEWIS[§]

Abstract. A growing literature studies social networks and their implications for economic outcomes. This paper highlights, examines, and addresses econometric problems that arise when a researcher studies network effects using sampled network data. In applied work, researchers generally construct networks from data collected from a partial sample of nodes. For example, in a village, a researcher asks a random subset of households to report their social contacts. Treating this sampled network as the true network of interest, the researcher constructs statistics to describe the network or specific nodes and employs these statistics in regression or generalized method-of-moments (GMM) analysis. This paper shows that even if nodes are selected randomly, partial sampling leads to non-classical measurement error and therefore bias in estimates of the regression coefficients or GMM parameters. The paper presents the first in-depth look at the impact of missing network data on the estimation of economic parameters. We provide analytic and numeric examples to illustrate the severity of the biases in common applications. We then develop two new strategies to correct such biases: a set of analytic corrections for commonly used network statistics and a two-step estimation procedure using graphical reconstruction. Graphical reconstruction uses the available (partial) network data to predict what the full network would have been and uses these predictions to mitigate the biases. We derive asymptotic theory that allows for each network in the data set to be generated by a different network formation model. Our analysis of the sampling problem as well as the proposed solutions are applied to rich network data of Banerjee, Chandrasekhar, Duflo, and Jackson (2013) from 75 villages in Karnataka, India.

JEL Classification Codes: C21, C49, C52, C83

Keywords: Social networks, Sampling of networks, Measurement error

Date: First Version: December 2008, This Version: November 21, 2016.

We thank Abhijit Banerjee, Victor Chernozhukov, Esther Duflo, Matthew O. Jackson, Anna Mikusheva, Whitney Newey, and Rob Townsend for encouragement and guidance. We are also grateful to Daron Acemoglu, Isaiah Andrews, Lori Beaman, Yann Bramoullé, Emily Breza, Denis Chetverikov, Albert Fang, Paul Goldsmith-Pinkham, Ben Golub, Jerry Hausman, Nathan Hendren, Guido Imbens, Anton Kolotilin, Horacio Larreguy, Ben Olken, Adam Sacarny, Paul Schrimpf, and Jan Zilinsky. Essential feedback was provided by seminar participants at Berkeley, Chicago Booth, Columbia, Harvard, HKS, Microsoft Research, Northwestern, NYU, Stanford, the Calvó-Armengol Prize Workshop 2010, NEUDC 2010, and various MIT workshops. Chandrasekhar is grateful for support from the NSF GRFP. This paper is based on chapters 1 and 2 of Chandrasekhar's dissertation.

[‡]Stanford University, Department of Economics and NBER.

[§]Netflix.

1. Introduction

A growing literature examines social networks and their implications for economic outcomes (see e.g., Jackson, 2008b, 2009a,b for an extensive survey of the literature). A network represents a set of connections (edges) among a collection of agents (nodes). For example, in a village network, nodes may represent households and edges may represent social ties between households. Applied researchers typically construct a network from data that has been collected from a partial sample of nodes rather than from all nodes in the network. Henceforth, such a network will be called the *sampled network*. This sampled network is often naively treated as the true network of interest. The researcher uses a collection of sampled networks to estimate how network structure affects economic outcomes. In our example, a researcher may study how the shape of the social network influences a social learning process. This paper highlights, examines, and addresses econometric problems that arise when a researcher studies these network effects using sampled network data.

Examples of network-based regressions in applied work include Kremer and Miguel (2007), who study the diffusion of deworming pill take-up, and Hochberg et al. (2007), who regress fund performance on measures of network importance of venture capital firms.[1] The applied work typically has low sampling rates (the share of nodes sampled), with a median of 42%, and 2/3 of the papers having a sampling rate below 51% (see Appendix F for details). Despite the prevalence of partial sampling, its implications for the estimation of economic parameters are rarely considered. A notable exception is Conley and Udry (2010) who study the diffusion of information among pineapple farmers in Ghana. Aware of the sampling problem, they conduct robustness exercises.

Our goal is to analyze the effect of using sampled network data on the estimation of parameters in network models of economic behavior given a fixed underlying network structure. Henceforth, we call these the "economic parameters" without meaning to suggest that network formation itself is not economic.[2] In general, we are interested in parameters in a generalized method of moments (GMM) model, motivated by theory, describing the behavior of nodes in a network. The biases in estimates of economic parameters have not yet been systematically dealt with. While GMM is a general framework, two common classes of models allow us to explicitly characterize biases and are easier to work with due to their linearity: regressions of economic outcomes on network characteristics and regressions of a node's outcomes on its

[1]There are numerous other examples. Kinnan and Townsend (2011) study whether whether households that are socially closer to credit sources smooth consumption better. Leider et al. (2009b) and Goeree et al. (2010) study the effect of social proximity between pairs on the offers made in dictator games. Alatas et al. (2016) and examines whether networks with better diffusion properties actually induced greater information spreading. De Giorgi et al. (2010) study how network neighbors' major choices affect a student's own major choice. Banerjee et al. (2013) study how network centrality of the initially informed influences the diffusion of microfinance. Cai et al. (2015) study how information about an insurance product spreads through a social network and look at heterogenous peer effects by cenrality of the peer.

[2]While parameters which describe the process by which networks are formed certainly are economic, we reserve "economic parameters" in our environment for parameters that describe a process that occurs on fixed networks.

network neighbors' outcomes. After characterizing the biases, we propose two strategies to correct such biases: a set of analytical corrections for commonly used network statistics and a two-step estimation procedure using graphical reconstruction – i.e., integrating over the missing data – that can be applied more broadly.

We focus on a running example throughout the paper: the diffusion of microfinance in 43 villages in rural Karnataka, India (Banerjee et al., 2013). A microfinance institution (MFI) based in Bengaluru expanded into these villages. Upon entering a village, the MFI informed certain households about its intentions and asked them to encourage others to join. The researcher wants to estimate how networks affect the diffusion of microfinance participation through these villages.

The present paper makes two core contributions. *Our first contribution is to highlight and analyze the biases in estimates of economic parameters when using sampled network data.* We develop analytical examples for commonly used network statistics, motivated by an applied economics questions concerning diffusion of information, social collateral, and risk-sharing. Next, we derive the corresponding biases that emerge when each of these statistics is used in regression. We show that even with random samples of the network the standard argument for attenuation due to classical measurement error does not apply; coefficients may expand, attenuate, or switch signs depending on the network statistic of interest. In addition, we consider a model in which a node's outcome depends on its peers' outcomes and a node's peer group is defined by the set of its social connections (Bramoulle et al., 2009; De Giorgi et al., 2010). We show that the instrumentation technique used in the literature to overcome the reflection problem (Manski, 1993) in such models is invalid since the measurement error in the instrument will be correlated with the measurement error in the endogenous variable. Similarly, we consider GMM estimation of the Jackson and Rogers (2007b) model of diffusion and show that sampling the network induces expansion bias in the diffusion parameter. We supplement our analysis with numerical evidence for a wide array of examples to illustrate how sensitive econometric estimation is to the sampling of a network. In our numerical experiments, we estimate many models across a number of network statistics. At a sampling rate of 1/3 we find that the estimates of the economic parameters have a mean absolute bias of 90% with a maximum of 260% for network-level regressions and a mean absolute bias of 63% with a maximum of 91% for node-level regressions.

Our second contribution is to develop two strategies to alleviate the biases: analytical corrections that apply to commonly used network statistics and two-step estimation using graphical reconstruction, which uses the observed part of the data to probabilistically reconstruct the missing part and then estimate the economic parameter accordingly.

First, by explicitly characterizing the biases, we derive simple bias corrections when the problem is tractable. We discuss several corrections and explore their reliability in addressing the biases. While computationally simple and easy to implement, these methods are typically limited to network-level regressions and are dependent on the particular network statistic of interest. Thus we develop a second, more general method that works well in practice – estimation by graphical reconstruction – to consistently estimate the economic parameter. Graphical reconstruction is simply asking the researcher to estimate the probability distribution governing each network and then integrate over the missing data. Integration over missing-at-random data is, of course, a well-practiced technique. We are particularly interested in the complications arising from the fact that researchers will have many different networks, each representing one draw from a distinct distribution, and ask whether, despite the extensive heterogeneity, researchers can consistently recover their economic parameters. This is analogous to fixed effects in a non-linear panel model. We establish checkable sufficient conditions under which the heterogeneity in the data can be respected and yet the economic parameter of interest is still consistently estimable. We check that several models satisfy these conditions. Note that graphical reconstruction does not limit the researcher to network-level regressions nor to specific and tractable network statistics.

Consider the case where a researcher wants to perform a network-level regression of the rate of microfinance participation in a village on the average path length of the network.[3] Without data on the entire network, the researcher falsely codes some existing links between individuals as if they do not exist. Graphical reconstruction builds on the simple idea that replacing every regressor for each village with a conditional expectation of the regressor delivers a consistent estimate of the regression coefficient. In our example, instead of using the mismeasured average path length of each network, the researcher ought to use the conditional expectation of the average path length, given the observed data. This requires integrating over all the missing data, given the observed information and sampling scheme, as opposed to treating missing links as if they did not exist. What complicates matters is that because different village networks form heterogeneously – potentially depending on different distributions – we are interested in the case where researchers respect this heterogeneity in their analysis. To do this, researchers need to accurately estimate the distribution governing each network's draw tightly and then integrate over the missing data. By treating every network as an independent, but not identically distributed, random variable, we estimate the conditional expectation of the average path length in every network and consistently estimate the regression coefficient.

In practice, the researcher will have to estimate the distribution of missing links. We propose a two-step procedure. In the first stage, the researcher fits a potentially different model of network formation to each network in the sample by making use of the observed data. We take no stand on the particular model and leave it to the researcher. Having done so, the researcher uses the network formation models to take draws of networks from their respective distributions, conditional on the observed information. Using these draws, the researcher estimates the conditional expectation of the regressor or moment in a GMM setting. In the second stage these conditional expectations are used in the usual way to

[3]The average path length is the mean of all shortest paths between all pairs of nodes in a network. Shorter paths mean that nodes are more likely to hear about information in most reasonable models.

estimate the economic parameter of interest. Conley and Udry (2010) perform a robustness exercise where they estimate missing neighborhood data in their regression model, which is an instance of graphical reconstruction.[4]

This two-step procedure is useful for several reasons. First, it allows the researcher to capture realistic heterogeneity by estimating a different model of network formation for every network.[5] Second, our theoretical frame is general, and we establish sufficient conditions ensuring that a desired class of network formation models can be used in graphical reconstruction. To build intuition, we make the analogy with panel data. Every network (individual in a panel) is independent, but the edges within a network (outcomes for an individual across time periods) exhibit dependence. Under regularity conditions, a large network, similar to a large time series, may contain enough information such that the researcher can use the observed data to accurately estimate the distribution which generates the network formation process (see, e.g., Chandrasekhar and Jackson (2016)). The technical challenge that we overcome is to control an incidental parameter problem, where a parameter for every network must be estimated.[6] Third, in our numerical experiments, it performs well. Even at a sampling rate of 1/3, the median bias is 5.7% for network-level regressions and 1.4% for node-level regressions. The median reduction in bias is 62%. Each of the 96 estimated parameters shows reduction in bias when the reconstruction estimator is applied. Fourth, in addition to regression of economic outcomes on network statistics, the methodology can be applied to GMM models and those with a family of moment functions indexed by some parameter which presents technical challenges. Covering these cases is essential to network analysis because natural models, such as stopping time models for diffusion, may carry an index.

Of course, given this procedure requires integrating over missing links, it is more demanding in terms of assumptions. In addition to having a collection of sampled networks and outcome variables, we assume that the researcher has covariates for each node (or pair of nodes) that will be predictive in the network formation models. Examples include GPS coordinates, ethnicity, and caste, which are often readily available in development applications and are obtained during the listing process in each enumeration area. In the economics of education settings, consider school networks, where it is straightforward to obtain school rosters and demographic data for the entire collection of students.

To demonstrate another practical application of our results, we describe how researchers can employ our framework to make better decisions in collecting sampled network data, given their budget constraints. We provide an algorithm to assess the trade-off between the number of networks in a sample and the sampling rate a researcher uses. This exercise is similar in

[4]The present paper develops a general theoretical framework, along with asymptotic analysis, that nests this strategy. We believe that estimates from graphical reconstruction ought to be used not only for robustness checks but also as estimates in their own right that exhibit substantially less bias. Additionally, with 4 networks in a sample, it is unlikely that the approach delievers consistent estimates in their case.

[5]In fact, we show in our empirical analysis that the degree of heterogeneity across our sample of 75 Indian village networks is such that treating them as draws from the same distribution can lead to misleading results.

[6]This resembles non-linear fixed effects in a panel (Hahn and Kuersteiner, 2004; Hahn and Newey, 2004).

spirit to power calculations frequently used in applied field work. First, the researcher obtains 100%-sampled network data for a small number of randomly chosen villages, using a pilot budget. Second, the researcher performs a numerical experiment by simulating outcome data from a specification that the researcher anticipates studying. In our microfinance example, the researcher simulates outcome data as a function of the path length from the initially informed households by assuming a regression coefficient and an R^2. Third, the researcher draws, with replacement, a set of villages sampled at a given rate such that her budget is exhausted. By applying graphical reconstruction, the researcher can assess the mean-squared error minimizing choice of sampling rate (or other sensible loss functions of their choosing).

We then apply our analysis of the sampling problem as well as the proposed solutions to sampled network data, collected in part by the authors, from 43 villages in Karnataka, India. Banerjee et al. (2013) study the diffusion of microfinance and, inspired by this analysis, we study natural specifications motivated by diffusion theory. We examine parameter estimates using the raw sampled data and compare them to those obtained by applying graphical reconstruction or analytical corrections. We find that applying our methods at times greatly changes parameter estimates and economic inferences. For instance, the impact of the network centrality of initially informed households on the microfinance take-up rate in the village is under-estimated by 33% using the raw sampled network data when compared to using graphical reconstruction. In addition, a regression of a node's take-up decision on its neighbors' decisions shows that endogenous network effects may be severely under-estimated (with a 60% bias relative to the corrected estimate) or even switch signs (with a 166% bias relative to the corrected estimate). Moreover, regression coefficients in several specifications are not significantly different from zero at conventional levels when using the raw sampled data but are significantly different when applying the reconstruction estimator.

Related literatures across a number of fields including economics, epidemiology, statistics, sociology, and computer science have extensively noted problems due to partial network data. The classical literature begins with Granovetter (1973), Frank (1980, 1981), and Snijders (1992) who identify how average degree and clustering are affected by several modes of random sampling. Rothenberg (1995) provides an excellent overview of the literature. More recently, the literature has focused on two classes of numerical experiments, typically with a single network.[7] The first class documents biases that emerge when estimating parameters in a network formation model with partial data (e.g., in economics, Santos and Barrett, 2008). Second, the literature numerically describes behavior of certain network statistics under sampling (e.g., in epidemiology, Ghani et al., 1998 and sociology, Kossinets, 2006). Handcock and Gile (2010) offer the straightforward solution to the first problem: by augmenting the likelihood to include the sampling scheme one can, in expectation, recover the correct network formation parameter.

[7] Santos and Barrett (2008) also provide an extensive discussion of survey methodology and Thompson (2006) discusses sampling methodology and inferences on the degree distribution and network size.

Our work builds on the above literature, with several key differences. First, prior to (the first version of) this paper, the literature typically has not focused on nor developed a methodology to consistently estimate parameters from models of behavior on networks with sampled data. The substantive distinction here must be stressed. We are not interested in recovering the structural properties of the unobserved part of the network *per se*; instead, our goal is to understand the biases in estimation of these economic parameters and develop a method to recover them.[8] Second, while augmented likelihood techniques for missing data are well-known in econometrics and statistics, we note that a collection of networks provides the researcher with a unique opportunity to set up the reconstruction problem in a manner which respects the substantial heterogeneity across networks. That is, a number of technical assumptions needed to control incidental parameter problems (e.g., nonlinear panel with fixed effects, Hahn and Kuersteiner, 2004; Hahn and Newey, 2004) become very palatable in the network context, given that each network carries within it tremendous amounts of information. Consequently, graphical reconstruction focuses on conditional expectations of network regressors or moments to consistently estimate economic parameters when graphs are drawn from heterogenous network formation models. This environment generates distinct technical challenges.

Finally, our work is of course related to the recent explosion in network formation models. An incomplete list of recent work includes Currarini et al. (2009), Christakis et al. (2010), Goldsmith-Pinkham and Imbens (2013), Boucher and Mourifié (2012), Leung (2014, 2013), Kolotilin (2013), Graham (2014), Mele (2016), Badev (2016), Sheng (2016), de Paula, Richards-Shubik, and Tamer (2014), and Menzel (2016), which all develop econometric network formation models. These works vary in aspiration and implication (e.g., partial versus point identified, consistently estimable parameters or not, nature of microfoundations), a discussion of which is well-beyond the scope of this paper. See de Paula (2015) for an excellent review. From the point of view of this paper, the researcher is interested in network formation models that can capture the attributes of the network that are relevant to her, and where parameters can be consistently estimable in a way that allows her to recover her economic parameters of interest. Note that modeling network formation, per se, is not the goal here.

The rest of the paper is organized as follows. Section 2 establishes the framework. The main results are in sections 3 and 4. Section 3 provides analytical examples of bias along with corrections. Section 4 discusses graphical reconstruction estimation. Section 5 contains numerical experiments which supplement sections 3 and 4. Section 6 applies the results to a study of the diffusion of microfinance. In section 7 we offer an algorithm for researchers to trade off the sampling rate against the number of networks. Section 8 concludes. All proofs are in the appendices.

[8]In that sense, while our conditional expectation of the structural properties of any given network will be accurate, perhaps we do not correctly recover the exact structure. However, we will consistently recover the economic parameter of interest without question.

2. Framework

2.1. **Notation and setup.** A network or a *graph* is a pair $G = (V, E)$ consisting of a set V of *nodes* and a set E of *edges*, with $n := |V|$. Nodes i and j are either connected or unconnected (the graph is unweighted) and if i is connected to j, then j is connected to i (the graph is undirected). Most of what follows in this paper is applicable to directed and weighted graphs, though following the bulk of the applied research we restrict our attention to the undirected, unweighted case. A graph with n nodes is a member of the set of all undirected, unweighted graphs, denoted by $\mathcal{G}_n$.

A graph is represented by its *adjacency matrix*, $A := A(G)$. It is a matrix of 0s and 1s that depicts whether two nodes are connected, where $A_{ij} = \mathbf{1}\{ij \in E\}$ with the convention that $A_{ii} = 0$. We denote the *neighborhood* of i, the set of nodes it is connected to, by $N_i := \{j \in V : ij \in E\}$. Researchers are interested in economic models where an economic behavior or outcome is predicted by network statistics. We let $w(G)$ represent a d_w-dimensional vector of these network statistics. Since the dataset may contain multiple networks, we use R to denote the number of graphs. The researcher is interested in *economic parameter* β_0.

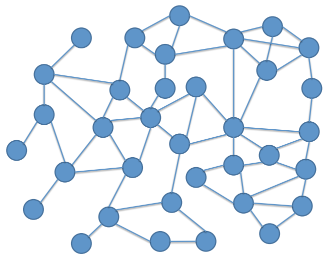
2.1.1. *Sampling.* We focus on the two most common types of sampled network data. First, the researcher may survey a set of m nodes and ask each node about the social connections with the other $m-1$ nodes in that data set. This is the *induced subgraph*, as it restricts the network among those who are sampled. Second, the researcher may have a list of the nodes in the network (e.g., a household census list in a village). From this list, a sample of m nodes may be surveyed. These nodes can name their social connections, not only to other $m-1$ surveyed nodes, but connections to anyone from the list of n nodes. This is the *star subgraph*, because the observed network is a union of stars with centers that are the sampled nodes.

Let S be the set of surveyed nodes, randomly chosen from V, with $m = |S|$. Let $\psi := \frac{m}{n}$ be the *sampling rate*. The researcher obtains a subgraph of the graph in question. There are two potential resulting networks: the induced subgraph $G^{|S} = (S, E^{|S})$, which consists of the sampled nodes and the edges restricted to the set of surveyed nodes ($E^{|S}$), and the star subgraph $G^S = (V, E^S)$, where E^S are edges such that at least one of the nodes is in S.

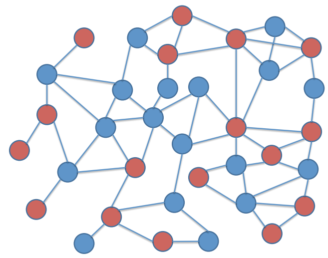
Figure 1 provides an illustration of the problem that this paper intends to address. Figure 1(a) displays G, the target network, Figure 1(c) shows the induced subgraph and Figure 1(f) depicts the star subgraph.

We also write $A = (A^{obs}, A^{mis})$ to denote the observed and missing part of the adjacency matrix, which are random variables under the sampling procedure. Although this framework idealizes the random sampling used in many applied contexts, our setting can easily be extended to other sampling methods such as independent edge sampling or snowball sampling.[9]

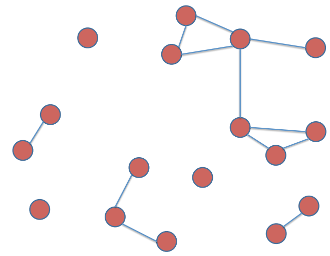
[9]Graphical reconstruction applies, with minimal modification of argument, to missing-at-random samples, where the probability of information being missing is independent of the missing data itself (Rubin, 1976).



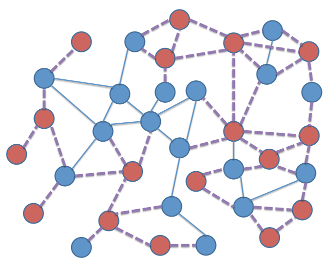
(A) Graph G



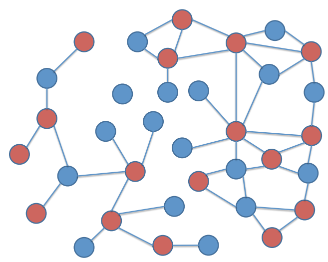
(B) Sampled nodes S



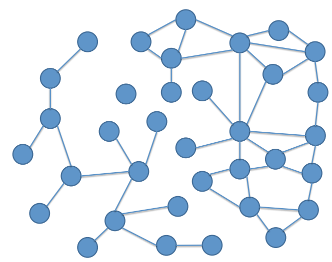
(C) Induced subgraph



(D) Highlights links used in star subgraph



(E) Star subgraph



(F) Star subgraph

FIGURE 1. Sampled nodes S in red. The figure depicts the induced subgraph and star subgraph.

2.2. **Econometric Models.** The researcher intends to study economic behavior on R networks, $\{G_r : r = 1, ..., R\}$.[10] For simplicity, we assume every network has n nodes. An economic process has taken place on every network and can be described by an econometric model depending on an *economic parameter* β_0. Returning to the microfinance example, information about microfinance has been introduced to certain households in every village and households decide to participate as the information propagates throughout the villages. Our goal is to estimate an economic parameter. We could easily do so if the networks were fully observed. The general framework for analyzing such models is to presume that a conditional moment restriction is satisfied,

$$\mathrm{E}\left[m(y, w(G); \beta_0)|G\right] = 0. \tag{2.1}$$

where $y \in \mathbb{R}^{d_y}$ is an outcome random variable, $m(\cdot, \cdot; \cdot)$ is a moment function, $w(\cdot)$ is function on $\mathcal{G}_n$, and $\beta \in \mathcal{B}$ is a parameter with true value β_0.

Examples include diffusion models (Banerjee et al., 2013; Alatas et al., 2016), discrete choice models, stopping time models (e.g., Iyer and Puri, 2012), quantile regression, and

[10] In this paper we take the view that the network is determined and then behavior operates on or through the network (e.g., social learning, peer effects, games on graphs) as compared to being interested in network formation itself.

network-based matching models (e.g., Aral and Walker, 2011).[11] Partial sampling will generally generate biases as the moment will be a nonlinear function of the network statistic, so the estimated parameter will be inconsistent.

While GMM is a general framework, two common classes of econometric models with network data are easier to analyze due to their linearity. The first class consists of models wherein economic outcomes are regressed on network characteristics. The second class consists of models where a node's outcome depends on its network neighbors' outcomes.

Class I: Regression of Economic Outcomes on Network Characteristics. Many researchers[12] study how network structure affects the economic outcome of interest, y, in regressions of the form[13]

$$y = \alpha + w\left(G\right)\beta_0 + \epsilon. \tag{2.2}$$

These researchers can estimate this regression at various observation levels. At the graph level, the data is $\{(y_r, w(G_r)) : r = 1, ..., R\}$ where $w(G_r)$ is a d_w-vector of network statistics (e.g., average degree, clustering, maximal eigenvalue, average path length) and the regression contains R observations. In our example, the researcher may regress the microfinance take-up rate in a village on the average network importance of the random set of households which were initially informed about microfinance. A simple model of diffusion in a network predicts that the centrality of these initial nodes correlates positively with take-up rates.

At the node level, the data is $\{(y_{ir}, w_{ir}(G_r)) : i = 1, ..., n,\ r = 1, ..., R\}$ where $w_{ir}(G_r)$ is a d_w-vector of statistics (e.g., degree of i, eigenvector centrality of i) and the regression has nR observations.[14] In our example, the researcher regresses a household's decision to join microfinance on its centrality. Theory suggests that central nodes will be more likely to learn new information. Similarly, one may estimate regressions at the edge level. Here $w_{ij}(G_r)$ is a d_w-vector of edge level statistics (e.g., social distance between the nodes) and the regression contains $\binom{n}{2} \cdot R$ observations.

Using sampled networks, the researcher runs regressions of the form

$$y = \alpha + w(\widetilde{G})\beta + u,$$

where $\widetilde{G}$ is either $G^{|S}$ or G^{S}, depending on the sampling scheme. In general, the measurement error will not be classical and may result in attenuation bias, expansion bias, or in pathological (in the univariate) or even mundane (in the multivariate) cases, sign switching. Sections 3.1 contain examples of common and economically meaningful network statistics where such biases exist and section 5 provides further numerical evidence on these biases.

[11] More generally, our results apply to indexed GMM models with parameter $\beta_0(u)$ where $u \in \mathcal{U}$ (e.g., time in a stopping time model or quantile in quantile regression), though we omit this.

[12] Examples include Leider et al. (2009b), Goeree et al. (2010), Cai et al. (2015), Banerjee et al. (2013).

[13] A vector of demographic covariates may be included, though we omit it for simplicity.

[14] With missing data, there are $O(nR)$ observations. For instance with $G^{|S}$, one has $mR = \psi nR$ observations.

Class II: Regression of Economic Outcomes on Neighbors' Outcomes. In a social equilibrium model, an economic outcome, y_i, depends on exogenous covariates of the individual, x_i, as well as the outcome of i's peer group, $\{y_j : j \in N_i\}$. In our running example, y_i is the microfinance meeting attendance rate of a household and x_i represents whether the researcher has exogenously informed the household. Estimating such a model is difficult in the usual way (Manski, 1993), but with network data, assuming exogeneity of x_i as in the above examples, Bramoulle et al. (2009) and De Giorgi et al. (2010) show that the model may be identified as the peer groups for individuals are overlapping but not identical.

Formally, let $y = (y_1, ..., y_n)'$ be the vector of outcome variables, $x = (x_1, ..., x_n)'$ be the vector of exogenous covariates and $\iota = (1, ..., 1)'$. A researcher is interested in estimating

$$y = \alpha_0 \iota + \rho_0 w(G) y + \gamma_0 x + \delta_0 w(G) x + \epsilon, \tag{2.3}$$

where $w(G)$ is a (possibly weighted) adjacency matrix that describes how much y_i is affected by others in the network. The economic parameter is $\beta_0 = (\rho_0, \gamma_0, \delta_0)$. Due to sampling, the researcher mistakenly estimates the model,

$$y = \alpha \iota + \rho w(\widetilde{G}) y + \gamma x + \delta w(\widetilde{G}) x + u, \tag{2.4}$$

where $\widetilde{w}$ is defined analogously with $\widetilde{G}$ either $G^{|S}$ or G^S. The neighborhoods will be misspecified and the estimator exhibits bias. We discuss this model in Section 3.2.

2.3. **Random Graphs and Asymptotic Framework.**

Random Graphs. Until now we have described an economic process, such as diffusion, occurring on a given collection of networks. Consider the example of a regression of y on network covariate $w(G)$. With missing data the researcher does not observe the true network statistic. In section 3 we demonstrate the biases induced by using $w(\widetilde{G})$ where $\widetilde{G}$ is the star or induced subgraph. Section 4 develops graphical reconstruction. We think of the network as the realization of a stochastic network formation process. We describe a number of examples in Section 4.4.1, but here we consider a simple but commonly used model: the probability that individuals i and j are connected, conditional on covariate z_{ij}, is given by

$$\mathrm{P}(A_{ij} = 1 | z_{ij}, \theta_0) = \Phi(z_{ij}' \theta_0),$$

where Φ is some link function. Thinking of the network as a random graph allows us to compute the conditional expectation of the regressor $w(G)$ given the observed portion of the network A^{obs} and the sampling scheme: $\mathrm{E}[w(G)|A^{obs}; \theta_0]$. If we knew the distribution of G we could compute this expectation. By properties of conditional expectation using $\mathrm{E}[w(G)|A^{obs}; \theta_0]$ as a regressor allows us to consistently estimate β_0.

Formally, each network G_r is a random graph is independently though not identically distributed over the space $\mathcal{G}_{n_R}$, where n_R is the number of nodes (which can depend on R for this thought experiment). We model the random networks as a triangular array of independent but not identically distributed random graphs, $G_{1,R}, ..., G_{R,R}$. Each $G_{r,R}$ is a random draw from a distribution $\mathrm{P}_{r,R}(G_r;\theta_{0r})$ over $\mathcal{G}_{n_R}$, where $\theta_{0r} \in \Theta_r$ is a parameter governing the distribution. We omit the R subscript indexing the triangular array.

Asymptotic Frame. Graphical reconstruction requires estimating a conditional expectation for every network. Since the parameter θ_{0r} for each network is unknown we must be able to consistently estimate all of these together. Intuitively, we need conditions such that every network has enough information in it so that its parameter can be precisely estimated. This is similar to panel data with non-linear fixed effects, where both the number of individuals and the number of periods grow.

Formally, we will assume that $n_R \to \infty$ as $R \to \infty$. The rate requirements of n and R are discussed in Section 4 and Appendix A.2. Moreover, $\Theta_{r,R}$ is typically finite dimensional, though we discuss an example where its dimension grows as $R \to \infty$. We assume conditions on n, R, and the random graph models such that every network asymptotically contains enough information to estimate θ_{0r} very well. In turn, we can estimate the conditional expectation very accurately and therefore recover the economic parameter β_0.

Finally, we employ the following notation throughout the paper. $\mathrm{E}[\cdot]$ denotes expectation, $\mathbb{E}_n[\cdot]$ the empirical expectation.[15] We will also make use of standard $O(n)$, $o(n)$, and $\Theta(n)$ notation. Note $f_n \in \Theta(g_n)$ means $\exists k_1, k_2 > 0, n_0$ such that $\forall n > n_0$ $|g_n| \cdot k_1 \leq |f_n| \leq |g_n| \cdot k_2$.

3. Analytical Examples of Bias

In this section we provide analytical examples which demonstrate the biases due sampled network data for three common classes of models: regression of economic outcomes on network statistics, regression of outcomes on network neighbors' outcomes, and a nonlinear GMM model of diffusion. The goal is to provide exact characterizations and develop intuitions for the sorts of biases under common forms of random sampling.

3.1. Regression of Economic Outcomes on Network Characteristics.

To gain intuition, it is useful to recall general measurement error in OLS. If the researcher is interested in a regression

$$y_r = w_r \beta_0 + \epsilon_r$$

but instead uses mismeasured regressors $\widetilde{w}_r$, the resulting estimator satisfies

$$\operatorname{plim} \widehat{\beta} = \beta_0 \frac{\operatorname{cov}(\widetilde{w}, w)}{\operatorname{var}(\widetilde{w})}.$$

Expansion, attenuation, and sign-switching bias are possible without any other assumptions.

In our environment, $w_r = w(G_r)$, the relevant network statistic, but due to sampling the researcher uses $\widetilde{w}_r = w(\widetilde{G}_r)$, where $\widetilde{G}_r$ is the star/induced subgraph. Thus, we are interested in the covariance of the network statistic with its true value, under the sampling scheme.

The covariance is typically not tractable to characterize. However, sometimes mismeasurement has a scaling effect in expectation. The scaling effect roughly means that $\mathrm{E}[\widetilde{w}|w] =$

[15] For $a = (a_1, ..., a_n)$, $\mathbb{E}_n[a_i] = \frac{1}{n}\sum_i a_i$. Similarly, $\mathbb{E}_R[a_r] = \frac{1}{R}\sum_r a_r$ and $\mathbb{E}_{n,R}[a_{i,r}] = \frac{1}{nR}\sum_r \sum_i a_{i,r}$.

$\pi w + o(1)$, where $\pi = \pi(\psi)$ is some known deterministic function. Clearly

$$\operatorname{plim} \widehat{\beta} = \beta_0 \cdot \underbrace{\pi^{-1}}_{\text{Scale}} \cdot \underbrace{\frac{\operatorname{var}(w)}{\operatorname{var}(w) + \operatorname{var}(v)\pi^{-2}}}_{\text{Classical attenuation}}$$

where $v = \widetilde{w} - \mathrm{E}\left[\widetilde{w}|w\right]$.

Note the two sources of biases: a scale effect which depends purely on $\pi(\psi)$, and can expand or attenuate the magnitude, and a dispersion effect which generates attenuation. Average degree and graph clustering are commonly used network statistics that exhibit scale transformations. However, more general statistics such as path length and eigenvalues are not merely scaled in this manner. Finally, it is easy to note, by the Cauchy-Schwarz inequality, that by looking at standarized effects, a researcher's conclusions are always conservative under sign-consistency.

Lemma **1.** *Let $\widetilde{w}$ be a mismeasurement of w with $\operatorname{cov}(\widetilde{w}, w) > 0$. Then $\operatorname{plim} \sigma_{\widetilde{w}} \widehat{\beta} \leq \sigma_w \beta_0$.*

As is well-known, the multivariate case is more complicated, though it is the more relevant case. Researchers often work with multivariate regression with network features (e.g., Cai et al. (2015)), but the story becomes messier. Consider a bivariate regression

$$y_r = w_r^1 \beta_0^1 + w_r^2 \beta_0^2 + \epsilon_r.$$

Let w_r^2 be observed without noise, w_r^1 be measured with noise, and $\operatorname{cov}(w^1, w^2) \neq 0$. It turns out that both estimates can be inconsistent and, moreover, the standardized effects need not even be conservative anymore. The result depends on a number of parameters including the variance of each regressor, their covariance, and parameters of the defining equation. (See Pischke (2000) or Hyslop and Imbens (2001) for details.)

To make this concrete for our setting, before continuing on to our analytic characterizations, we take a simple, bivariate example of two network features: the *average degree* and *clustering*. The average degree of a network, $d(G) := n^{-1} \sum_i \sum_j A_{ij}$, is the average number of links each node has in the network. The *average clustering* is the average, over all nodes in the network, of the share of a node's neighbors that are themselves linked.

Figure 2 presents results from a simulation where we assume average degree is not measured with error, whereas average clustering is due to sampling.[16] Here we vary the correlation between clustering and average degree in the graphs that we draw as we also vary the sampling rate. We can see that $\widehat{\beta}^1$ (the coefficient on clustering) can expand or attenuate but, moreover, $\widehat{\beta}^2$ can expand, attenuate, and switch signs. This happens even though the regressor (degree) was not measured with error.

In what follows, we return to looking at univariate regressions or simple linear-in-means or GMM models geared to highlighting intuitions of the bias, but we caution that the biases in the multivariate cases are considerably more complicated.

[16]The idea is that under star sampling, as seen below, it is easy to not mismeasure average degree. We assume the researcher has not corrected for mismeasurement in clustering.

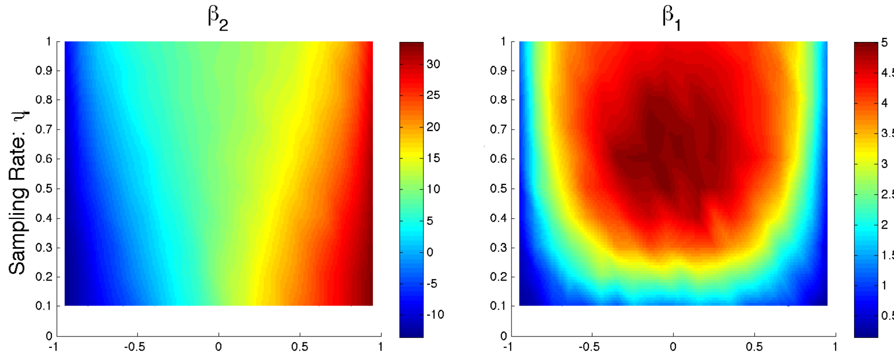


FIGURE 2. Sampling rate ψ against correlation between average clustering (w^1) and average degree (w^2). Heat map denotes estimate $\widehat{\beta}^j$ for $j \in \{1, 2\}$. Note that $\beta_0^1 = 2.5$ and $\beta_0^2 = 10$.

Local Statistics. We begin with a warm-up and look at local network statistics. Here we mean that the features of the statistic only depend on a node's immediate neighbors or, possibly a node's neighbors and their common edges. The statistics we look at are the average degree (average number of links), the graph clustering (how likely neighbors' neighbors are linked), and support (how likely there is some node that is linked to both nodes in an edge). Support is particularly important in models of risk-sharing and favor exchange (Jackson et al., 2012).

Recall average degree is the average number of links per node

$$d(G) = \frac{1}{n} \sum_i \sum_j A_{ij},$$

graph clustering is the share of two-stars that are triangles,

$$c(G) = \frac{\sum_{i<j<k} A_{ij} A_{ik} A_{jk}}{\sum_{i<j<k} 1\{A_{ij} + A_{jk} + A_{ik} = 2\}},$$

and support is the share of links whose nodes have at least one friend in common

$$s(G) = \frac{\sum_{i<j} A_{ij} 1\{\sum_k A_{ik} A_{jk} > 0\}}{\sum_{i<j} A_{ij}}.$$

Finally it is useful to define $\Pi(G) = 1 - \sum_x (1-\psi)^x \, \mathrm{P}(x|G)$, where $\mathrm{P}(x|G)$ is the share of links ij that are supported by exactly x nodes in the underlying graph G.

PROPOSITION 3.1. *Consider a regression $y_r = \alpha + w(G_r)\beta_0 + \epsilon_r$ with w being either average degree, graph clustering or support. Assume the data $(y_{rR}, w(G_{rR}))$ is a triangular array satisfying the regularity conditions of Assumption A.1.*

(1) *For average degree,*
$\widehat{\beta}(G^{|S}) \xrightarrow{\mathrm{P}} \frac{\beta_0}{\psi} \cdot \frac{\mathrm{var}(w)}{\mathrm{var}(w)+\psi^{-2}\mathrm{var}(v^{|S})}$ *and* $\widehat{\beta}(G^{S}) \xrightarrow{\mathrm{P}} \frac{\beta_0}{1-(1-\psi)^2} \cdot \frac{\mathrm{var}(w)}{\mathrm{var}(w)+(\psi(2-\psi))^{-2}\mathrm{var}(v^{S})}.$

(2) *For graph clustering,*
$\widehat{\beta}(G^{|S}) \xrightarrow{\mathrm{P}} \beta_0 \cdot \frac{\mathrm{var}(w)}{\mathrm{var}(w)+\mathrm{var}(v^{|S})}$ *and* $\widehat{\beta}(G^{S}) \xrightarrow{\mathrm{P}} \left(\frac{\psi(3-2\psi)}{1+\psi(1-\psi)}\right)^{-1} \beta_0 \cdot \frac{\mathrm{var}(w)}{\mathrm{var}(w)+\left(\frac{\psi(3-2\psi)}{1+\psi(1-\psi)}\right)^{-2}\mathrm{var}(v^{S})}.$

(3) *For support,*
$\widehat{\beta}(G^{|S}) \xrightarrow{\mathrm{P}} \frac{\beta_0}{\Pi} \cdot \frac{\mathrm{var}(w)}{\mathrm{var}(w)+\Pi^{-2}\mathrm{var}\left(v^{|S}\right)}$ *and* $\widehat{\beta}(G^{S}) \xrightarrow{\mathrm{P}} \left(\frac{\psi^2+2\psi(1-\psi)\Pi}{\psi^2+2\psi(1-\psi)}\right)^{-1} \beta_0 \cdot \frac{\mathrm{var}(w)}{\mathrm{var}(w)+\left(\frac{\psi^2+2\psi(1-\psi)\Pi}{\psi^2+2\psi(1-\psi)}\right)^{-2}\mathrm{var}\left(v^{|S}\right)},$
where $\Pi = \lim_R \frac{1}{R}\sum_r \Pi\left(G_{rR}\right)$.

These statistics are in increasing order of complexity. Note that degree only depends upon the immediate edges for each individual while clustering only depends upon the observability of common edges among a nodes' neighbors. Support is a nonlinear function that expresses the existence of connectivity among my neighbors and, hence, relies upon the graph structure, making elimination of the bias much more difficult.

It is useful to note that under simple assumptions $\mathrm{var}(v^S)$ and $var(v^{|S})$ are going to be $o(1)$. And in these cases, one only needs to just rescale estimates appropriately: in a trivial way in the case of degree, and in a more complicated way that involves estimating the distribution of supporting nodes in the case of support. Even if one is not in this case, it is easy to just estimate $\widehat{\mathrm{var}}(v)$ and directly use this to eliminate the bias.

By characterizing the bias, corrections are simple. Faced with induced subgraph sampling,

(1) for degree, rescale the estimate by the sampling rate,
(2) for clustering, use the estimate directly,
(3) but for support, there is not a clear simple solution.

Facing star subgraph sampling,

(1) for degree use only $i \in S$ or with the full sample transform the estimate by $1-(1-\psi)^2$,
(2) for clustering use only induced subgraph or with the full sample transform the estimate by $\left(\frac{\psi(3-2\psi)}{1+\psi(1-\psi)}\right)$.
(3) and for support use only $i, j \in S$ or use the set of $i, j \in S$ to estimate the distribution of the number of nodes that provide support and then explicitly correct for the bias using the full sample.

Our simulations suggest that one is better off using the full sample and estimating and eliminating bias, rather than restricting to just the sampled nodes or sampled pairs in the star subgraph case.

Global Statistics. Next we turn to what we call global network statistics. This is a catch-all term that refers to functions of arbitrary collections of edges. Thus, this includes interactions of individuals greater than distance two. Not surprisingly, studying biases in these cases becomes considerably harder and at times requires more approximation.

Path Length and Graph Span. The first example is the path length. The path length between two nodes i and j is given by the minimum number of steps taken on the graph to get from i to j, denoted $\gamma(i,j) := \min_{l\in\mathbb{N}\cup\infty}[A^k]_{ij} > 0$. If there is no such finite path, we put $\gamma(i,j) = \infty$.

The average path length of a graph is the mean taken over all finite paths,

$$\gamma(G) := \sum_{i,j:\gamma(i,j)<\infty} \gamma(i,j) / \left| \{(i,j) \in V^2 : \gamma(i,j) < \infty\} \right|.$$

Models of diffusion of information, flows of finance, risk-sharing, nepotism, and other phenomena, build on the principle that the farther apart agents are, the less is transmitted between them. For example, Kinnan and Townsend (2011) study how the network distance to a bank affects consumption smoothing. Other papers that use path length or average path length include Golub and Jackson (2010) who simulate diffusion processes; Leider et al. (2009b) and Goeree et al. (2010) who study dictator games between members of a school; Alatas et al. (2016) who look at the diffusion of information about poverty; and Banerjee et al. (2013) who study the diffusion of microfinance.

The basic idea, as seen from Figure 1 is that $\gamma_{\widetilde{G}}(i,j) \geq \gamma_G(i,j)$ and therefore paths are longer. However, the average path length is well-known to be a very difficult object to study analytically.[17] Both the economics and statistical physics literatures study an object we term the *graph span*, mimicking average path length. Jackson (2008a) shows that for a general family of random graph models the ratio of the graph span to average path length asymptotically almost surely is one. The statistical physics literature uses such an approximation as well (e.g., Newman et al., 2001; Watts and Strogatz, 1998, Watts and Strogatz, 1998). This motivates the study of the graph span as a regressor. Let $d_2(G) := n^{-1} \sum_{i=1}^n \sum_{j>i}^n \sum_{k \neq i,j} A_{ij} A_{jk}$ be the average number of second neighbors.[18] The *graph span* is

$$\ell(G) := \frac{\log n - \log d(G)}{\log d_2(G) - \log d(G)} + 1.$$

Larger networks have higher spans. Networks that are more expansive in the sense that the number of second neighbors far exceeds the number of neighbors have lower spans; it takes fewer steps to walk across the network. It is useful to define a constant which is a bound on the ratio of the size of a neighborhood to the size of a neighborhod's neighborhood: $c := \sup_{R \geq 1} \sup_{r \leq R} d(G_{rR})/d_2(G_{rR})$. Finally let $k(\psi) = \psi + \psi^2 - \psi^3$.

Proposition 3.2. *Consider a regression $y_r = \alpha + w(G_r)\beta_0 + \epsilon_r$ with w being the graph span. Assume the data $(y_{rR}, w(G_{rR}))$ is a triangular array satisfying the regularity conditions of Assumptions A.1 and A.2. Then,*

(1) *$\widehat{\beta}$ is sign-consistent with attenuation if $\psi \in (c,1)$ or $k(\psi)/(1-(1-\psi)^2) \in (c,1)$:*

$$\text{plim}\,|\widehat{\beta}(G^{|S})| < |\beta_0| \text{ and } \text{plim}\,|\widehat{\beta}(G^S)| < |\beta_0|.$$

[17] Bollobas (2001) approaches path length from an exact analytical perspective but only for a very specific random graph family. This approach is not suitable for gaining intuition for broader classes of graphs.

[18] Notice this defines second neighbor in the sense of taking a random node and then counting the number of neighbors of each of the node's neighbors. The definition is different from counting the number of distinct nodes at path length two from a given node, which would be $\frac{1}{n} \sum_i \sum_{k>i} \sum_{j \neq i,k} A_{ij} A_{jk} (1 - A_{ik})$.

(2) *and $\widehat{\beta}$ may be sign-inconsistent otherwise.*

Sampling a network thins out the set of edges, resulting in a higher graph span. As the graph span approximates behavior of average path length, it captures the intuition that due to sampling, paths on graphs seem longer than they truly are. The expansion of the graph span has a slope effect on $\widehat{\beta}$, and as $\log\psi < 0$ and $\log(k(\psi)/(1-(1-\psi)^2)) < 0$, the effect is either attenuation unless the sampling rate is too low, in which case sign-switching becomes a possibility. One must proceed with caution when discussing cases where the sampling probability is too low. In this case the network can shatter, yielding "islands" of disconnected sets of nodes which have short average path lengths within the set but have infinite distance across the sets.[19] Since average path length is defined as a mean conditional on all finite paths, this is precisely where sign-switching may occur in practice. Alatas et al. (2016) contains an example where this happens in Indonesian networks.

Having characterizing the biases, the solutions are simple.

(1) For induced graph sampling use

$$\widetilde{\ell}(G^{|S}) := \frac{\log(\psi^{-1}m) - \log(\psi^{-1}d(G^{|S}))}{\log(\psi^{-2}d_2(G^{|S})) - \log(\psi^{-1}d(G^{|S}))} + 1$$

(2) and for star graph sampling use

$$\widetilde{\ell}(G^{S}) := \frac{\log n - \log(m^{-1}\sum_{i\in S}\sum_j A(G^S)_{ij})}{\log(d_2(G^S)/k(\psi)) - \log(m^{-1}\sum_{i\in S}\sum_j A(G^S)_{ij})} + 1.$$

The corrected estimates are consistent as noted in the proof of Proposition 3.2.

Spectral Functions. Spectral functions are network statistics that relate to the set of eigenvalues of matrices which represent the graph, such as the adjacency matrix. They are useful in characterizing properties of the network. The distribution of eigenvalues has applications to models of information diffusion and risk-sharing as well. The number of k-length walks that cycle back to the original node correspond to k-th moment of the eigenvalue distribution, denoted $\mu_k(G)$,

$$\mu_k(G) = n^{-1}\sum_{i_1,..,i_k\in V^k} A_{i_1i_2}...A_{i_ki_1} = n^{-1}\mathrm{Tr}(A^k)$$

where $V^k = V\times ... \times V$ (Barabasi and Albert, 1999). Given that the graph spectrum carries a great deal of information about the diffusive properties of a network, it is a useful regressor.

There are several applications of spectral statistics in economic theory. For instance, the first eigenvalue of the adjacency matrix, $\lambda_1(G)$, describes how well the graph diffuses information (e.g., Bollobás et al., 2010).[20] In models of social learning Golub and Jackson

[19] One can check that a graph H with $d_2(H)/d(H) < 1$ cannot be connected. The sign-switching case requires at least some $d_2/d < 1$ which we note the researcher can immediately detect.

[20] In a percolation process the threshold probability above which a giant component emerges is precisely $1/\lambda_1$. For another intuition, if A is diagonalizable, then the dominant factor in $\left\|A^k\right\|$ is λ_1^k.

(2009, 2010) show that the second eigenvalue of a weighted adjacency matrix is related to the time it takes to reach consensus; similar results are shown in DeMarzo et al. (2003). Ambrus et al. (2010) also characterize the risk-sharing capacity of a network as a function of the expansiveness of the network; it is well-known in network theory that this maps into the eigenvalues of a transformation of the adjacency matrix (Chung, 1997). It is difficult to precisely characterize the behavior of these spectral regressors, though we present bounds on their behavior under sampling.

Proposition 3.3. *For an arbitrary graph G, we have*

(1) $\lambda_1(G^{|S}) \leq \lambda_1(G^S) \leq \lambda_1(G)$.
(2) $\mu_k(G^S) < \mu_k(G)$.
(3) $\mathrm{E}\left[\mu_k(G^{|S})|G\right] = \sum_{j=2}^{k} \frac{(m-1)_j}{(n-1)_j}\eta_j < \mu_k(G)$, *where η_j is the number of sets of j-distinct nodes that are counted.*

Since λ_1 can be thought of as measuring the number of walks through the graph (and with missing edges there are fewer walks), we expect expansion bias in $\widehat{\beta}$ when using these regressors. This follows from the interlacing theorem.[21] This means that networks will appear to be more diffusive than they actually are.

3.2. **Regression of Outcomes on Network Neighbors' Outcomes.** We discuss the impact of sampled networks on regressions of nodes' outcomes on network neighbors' outcomes. The models we consider are developed in Bramoulle et al. (2009) and De Giorgi et al. (2010) and naturally extend the models discussed in Manski (1993) to a network setting. Blume et al. (2011) contains an extensive review of the literature. The network allows for nodes to have overlapping but not identical peer groups.

The model is given by (2.3), and we are interested in $\beta_0 = (\rho_0, \gamma_0, \delta_0)$. There are two natural examples for how neighbors' outcomes ought to affect a node's outcome. First, every node's outcome may be affected by the average outcome of its neighbors.[22] Second, every node's outcome may be affected by the total sum of its neighbors' outcomes.[23] The reduced form is

$$y = \alpha\iota/(1-\rho_0) + \gamma_0 x + (\gamma\rho_0 + \delta_0)\sum_{k=0}^{\infty} \rho_0^k w^{k+1} x + \sum_{k=0}^{\infty} \rho_0^k w^{k+1}\epsilon.$$

Since a node's neighborhood outcome, wy, is the endogenous regressor, the reduced form suggests that extended neighborhood effects, powers $w^k x$ ($k \geq 2$), can be used as instruments

[21]Whether there is expansion bias depends on how the eigenvalues shrink across the initial distribution. For instance, if the contraction is by translation, the regression slope would clearly not change. Numerical with simulated networks and empirical data provide evidence of expansion bias.

[22]We can write the model as $y_i = \alpha + \beta\mathbb{E}_{N_i}[y_j] + \gamma x_i + \delta\mathbb{E}_{N_i}[x_j] + \epsilon_i$ as $\mathbb{E}_{N_i}[y_j] = \sum_{j\in N_i} y_j/d_i = \sum_j y_j A_{ij}/d_i$.

[23]We discuss the first case, though clearly by mimicking the argument the results follow for the second.

for wy. We focus on the instrument $Z = [\iota, x, wx, w^2x]$.[24] Setting $X = [\iota, wy, x, wx]$, the estimator is $(X'P_ZX)^{-1}X'P_Zy$.

Identification comes from intransitive triads.[25] If i and j are connected and j and k are connected, but i and k are not connected, then k affects i only through j. As such, x_k is used as an instrument for y_j's effect on y_i. We caution that this identification strategy convincingly works only when x is randomly assigned (e.g., Ngatia (2011)) as identification crucially depends on exogeneity of x.

We examine the estimation of (2.4) using $\widetilde{w} = w(G^S)$ or $\widetilde{w} = w(G^{|S})$ with instrument $Z_{\widetilde{G}} = [\iota, x, \widetilde{w}x, \widetilde{w}^2x]$. We show that the exclusion restriction is invalid when using sampled network data, even if the covariates are exogenous and the usual identification requirements are met if the full network data was available.[26]

Proposition 3.4. *Assume $\gamma_0\rho_0 + \delta_0 \neq 0$ and $w^2 \neq 0$, so 2SLS is valid for* (2.3). *Then 2SLS with*

(1) *$w(G^S)$ with (y_i, x_i) for all $i \in V$ generically yields* $\mathrm{E}\left[Z_{G^S}u_{G^S}\right] \neq 0$,[27]
(2) *$w(G^{|S})$ with (y_i, x_i) for all $i \in S$ generically yields* $\mathrm{E}\left[Z_{G^{|S}}u_{G^{|S}}\right] \neq 0$,
(3) *$w(G^S)$ with (y_i, x_i) for all $i \in V$ but restricting the second stage to $i \in S$, yields* $\mathrm{E}\left[Z_{G^S}u_{G^S}\right] = 0$.

Sampling induces an errors-in-variables problem, wherein the neighborhood effect is mismeasured since the neighborhoods themselves are misspecified. Though typically one uses instruments to address such a problem, here the instrument is correlated with the measurement error in the regressor, as the instrument involves powers of the mismeasured adjacency matrix. As such, the exclusion restriction is violated.

Figure 3 provides two examples where invalid instrumentation is generated. Figure 3(a,b) show that if j is sampled but i and k are not, the sampled network falsely suggests that k is a valid instrument for j's effect on i. Similarly, figure 3(c,d) show a case with the induced subgraph, where k instrumenting for j's effect on i will be invalid as the other channels through which k affects i are not accounted for due to sampling. In this case, the channel through l is omitted.

With G^S data, however, there is a simple analytical correction. For $i \in S$, notice that $[\widetilde{w}x]_i = [wx]_i$ and $[\widetilde{w}y]_i = [wy]_i$. Consequently, there is no measurement error in the second stage for these observations. As only the first stage contains measurement error, uncorrelated with the second stage residual, such an exercise satisfies the exclusion restriction.

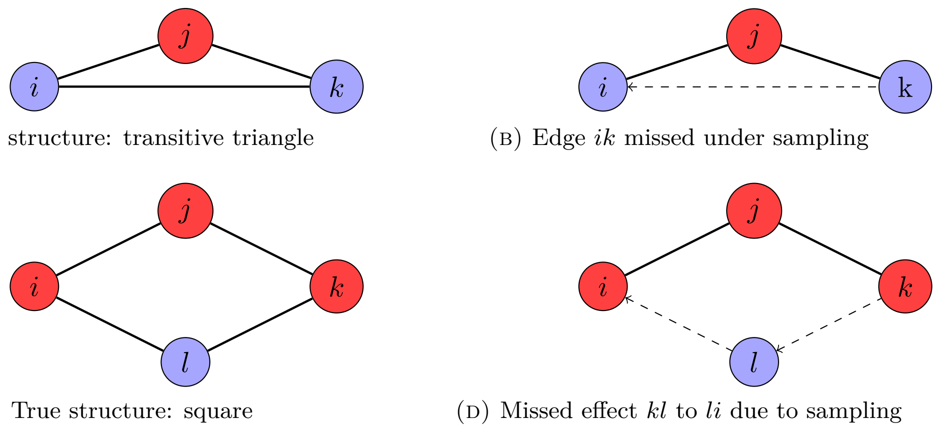
3.3. **A Model of Diffusion.** Having discussed several examples of network-based regressions, we now turn to a model of diffusion examined in Jackson and Rogers (2007b) which

[24]Other estimation strategies are suggested in the literature, on the basis of efficiency (Bramoulle et al. (2009); Lee et al. (2010)). They require the validity of the instrument Z.

[25]Bramoulle et al. (2009) provide formal identification conditions.

[26]De Giorgi et al. (2010) are aware that measurement error may cause problems in this model and conduct a numerical robustness exercise.

[27]We say generically in the sense that given (G, x, β_0), only a finite set of $\psi \in [0, 1]$ satisfy $\mathrm{E}\left[Z_{G^S}u_{G^S}\right] = 0$.



(A) True structure: transitive triangle

(B) Edge ik missed under sampling

(C) True structure: square

(D) Missed effect kl to li due to sampling

FIGURE 3. Red nodes are sampled. (a) and (c) show examples of true network structures. (b) shows the star subgraph from (a) when j is sampled. The effect of k on i is missed. (d) shows the induced subgraph from (c) when i, j, and k are sampled. The effect of k on i through l is missed.

we discuss in the context of our microfinance example.[28] The researcher is interested in estimating this diffusion model which satisfies equilibrium moment equations. There are two states: whether or not a household endorses microfinance in a weekly village gathering. A non-endorsing household with d_i links may choose to endorse with probability $\nu_0 d_i \sigma_i$ where ν_0 is a transmission parameter and σ_i is the fraction of i's neighbors that have decided to endorse. However, an endorsing household may naturally decide not to endorse, which can happen with probability δ_0. Jackson and Yariv (2007) extend this model to a number of strategic environments.

The model is identified up to parameter $\beta_0 := \nu_0/\delta_0$, which is the transmission to recovery rate. Let $P(d)$ denote the degree distribution and $\rho(d)$ the share of nodes with degree d that endorse. Finally, $\bar{\rho}_* := \sum_d \rho(d)P(d)$ is the average endorsement rate in the network and the researcher observes $y := \bar{\rho} + \epsilon$, with ϵ an exogenous zero mean shock.

The second neighbors endorsement rate is given by $\sigma = (\mathrm{E}d)^{-1} \sum_d \rho(d)P(d) \cdot d$. Jackson and Rogers (2007b) use a mean-field approximation to derive a steady state equation,

$$\rho(d) = \frac{\beta_0 \sigma d}{1 + \beta_0 \sigma d}.$$

The equilibrium satisfies

$$\sigma(\beta_0) = (\mathrm{E}d)^{-1} \sum_d \frac{\beta_0 \sigma(\beta_0) d^2}{1 + \beta_0 \sigma(\beta_0) d} P(d). \tag{3.1}$$

[28] This class of models, developed by Pastor-Satorras and Vespignani (2001), have been extended to numerous strategic interaction settings by Galeotti and Rogers (2013) and Jackson and Yariv (2007), who study the SIS (susceptible, infected, susceptible) model of epidemiology, which they and others show have applications in a wide variety of economic contexts.

By combining (3.1) with the definition of $\bar{\rho}$, we find that $h(G_r;\beta) := \sum_d \frac{\beta\sigma_r(\beta)d}{1+\beta\sigma_r(\beta)d}P(d) = \bar{\rho}$. Therefore the researcher can use as moments

$$m(y_r, G_r;\beta) := y_r - h(G_r;\beta),$$

and estimate β_0 via nonlinear least squares. Jackson and Rogers (2007b) show that an equilibrium with non-zero endorsement rate exists only if $\beta > \mathrm{E}d/\mathrm{E}d^2$. The ratio of squared degree to degree, similar to what we have encountered when studying graph span, again becomes an important feature of the network. We put $\zeta := \mathrm{E}d^2/\mathrm{E}d$. Note that the typical summand of $h(G_r,\beta)$ is monotone and convex in d. Therefore, stochastic dominance relations among various distributions $P(d)$ will play a central role.

As one expects, due to sampling of networks the researcher will overestimate the transmission parameter. An intuition for this is provided by the case of the star subgraph. This form of subsampling leads to a degree distribution that will be first order stochastically dominated by the true distribution. Therefore, the sampled network seems as if it has poorer diffusive properties; to generate the same average endorsement rate, the parameter governing the diffusion process must be higher. In addition, we show that the diffusion with the true parameter β_0 occurring on the sampled network may have no non-zero equilibria. When β_0 is close enough to the threshold $1/\zeta(G)$, the partially sampled network will make threshold ratio $1/\zeta(\widetilde{G})$ rise and therefore β_0 may appear to be less than $1/\zeta(\widetilde{G})$.

Proposition 3.5. *Assume we have a triangular array (y_{Rr}, G_{Rr}) with degree distributions $\mathrm{P}_{Rr}(d)$ and (i) (3.1) holds in expectation for each r, (ii) β_0 is such that there is a positive endorsement in every equilibrium, (iii) $\mathcal{B}$ is a compact subset of $\mathbb{R}_{++}$, (iv) (ϵ_r) are iid zero mean finite variance disturbances, and (v) $\limsup_{R\to\infty}\sup_{r\le R}\sup_d |\mathrm{P}_{Rr}(d) - \mathrm{P}_{\infty r}(d)| = 0$.*

(1) *The estimates exhibit expansion bias:* $\operatorname{plim}\widehat{\beta}(G^S) > \beta_0$ *and* $\operatorname{plim}\widehat{\beta}(G^{|S}) > \beta_0$.
(2) *For all r, β_0 is outside the range generating positive endorsement rate in the estimated equilibrium, with probability approaching one, under the following additional assumptions. Put $\delta_r := \beta_0 - 1/\zeta_r > 0$ and assume*
   (a) *for star subgraphs:*
      - $\liminf_{R\to\infty}\zeta_r > 1+\psi$
      - $\limsup_{R\to\infty}\delta_r < (1-\psi)\cdot\frac{1-\zeta_r^{-1}(1+\psi)}{\zeta_r+(1-\psi^2)}$

   (b) *or for induced subgraphs,* $\limsup_{R\to\infty}\delta_r < (1-\psi)\cdot\frac{1+\zeta_r^{-1}}{\psi\zeta_r+(1-\psi)}$.

It is easy to see that for the star subgraph, an analytical solution to the bias is to use the degree distribution of the sampled nodes. However, this is a highly non-generic solution. The induced subgraph, for instance, does not allow this approach nor do other sampling schemes (e.g., randomly chosen edges, etc.). A natural question to ask is whether we may use the sampled degree distribution, such as $P^{|S}(d)$, to obtain $P(d)$. We note that this will not be straightforward to do, in general, because it generates an ill-posed inverse problem. The researcher is faced with an under-determined system; while we can describe how $P(d)$ maps into $P^{|S}(d)$ due to sampling, there appears to be no unique inverse. Graphical reconstruction, however, will provide a way to address the problem.

3.4. **From Analytical Examples to Graphical Reconstruction.** In this section we have analytically examined biases that emerge from sampled networks. We focused on network statistics motivated by a number of applied questions concerning diffusion of information, network importance, risk-sharing, and social collateral. By analytically characterizing the biases, we were able to describe the mechanics of the non-classical measurement error and construct analytical corrections to eliminate the biases, under regularity conditions. The analytical study required us to focus on graph-level regressions; moreover, to be consistent, the analytical corrections focused on eliminating a slope effect of the biases, but needed to assume away or estimate a dispersion effect.

We also examined a model where a node's outcome depends on its neighbors' outcomes and demonstrated that a network-based instrumentation method violates the exclusion restriction when the network is sampled. With certain data structures, we provided a simple solution. Furthermore, we extended our analysis to a GMM model of diffusion and pointed out how the estimated parameters would exhibit expansion bias.

In general, our discussion has been on a case-by-case basis in this section. We have mostly focused on graph-level regressions and have been only able to examine very tractable network statistics. Numerous network statistics such as betweenness centrality, eigenvector centrality, and the aforementioned spectral statistics do not permit easy analytical examination nor corrections. The next section provides a more general method to estimate the economic parameter. Though the method is not limited to graph level regressions nor tractable network statistics, it comes at the cost of requiring more data and putting more structure on the problem by assuming models.

4. Graphical Reconstruction

In this section, we discuss a two-step estimation procedure to consistently estimate economic parameters from linear regression and GMM models. In our asymptotic frame, both the size of each network and the number of networks grow. Every network is a draw from a distribution governed by its own parameter θ_{0r}. This will force us to control an incidental parameter problem. Clearly, we can nest the special case where every network is drawn from the same distribution, $\theta_{0r} = \theta_0$ for every r, and thereby assume away the incidental parameter problem. However, based on our experiences with empirical data, forcing every network to be drawn from the same model introduces enough misspecification to negate the benefits of graphical reconstruction.

We present an informal overview in section 4.1. In section 4.2, we present the asymptotic distribution of $\widehat{\beta}$ under high level assumptions on $\widehat{\theta}_r$, the regularity conditions for which are listed in Appendix A.2, and detail the estimation procedure in section 4.3. We provide sufficient conditions for network formation models that allow for $\widehat{\theta}_r$ to satisfy the aforementioned high-level conditions and check several classes of network formation models, which also shed light on the limits of our approach, in section 4.4.

4.1. **Informal Overview.** In our overview we describe our procedure for regression,

$$y_{ir} = \alpha_0 + w_{ir}(G_r)\beta_0 + \epsilon_{ir}.$$

We assume that the researcher has the following data. First, she has outcome data for every node in every graph, $\{y_{ir} : i = 1, ..., n,\ r = 1, ..., R\}$, such as whether household i in village r participates in microfinance.[29] Second, she has a set of partially observed graphs, $\{G_r^S : r = 1, ..., R\}$ or $\{G_r^{|S} : r = 1, ..., R\}$. Third, she has variables which are predictive in a network formation model $\{z_r : r = 1, ..., r\}$.[30] For instance, the researcher may have basic demographic characteristics such as religion, caste, household amenities, occupation or geographic location. This data structure is relatively innocuous and common in numerous applications. In development economics, when deciding how to draw a random sample to administer treatments, researchers usually conduct a listing in each enumeration area. This requires obtaining a census of the economic units, which can be done directly (e.g., Townsend, 2007; Suri (2011); Banerjee et al., 2013) or indirectly by obtaining census information from the village representatives (e.g., Macours, 2003; Takasaki et al., 2000; Alatas et al., 2016).[31] It is well-known that obtaining GPS and basic demographic data during enumeration is cheap; the bulk cost of a network survey is the network module itself. For a different example, consider school networks where it is straightforward to obtain rosters and demographic data for all students. The full set of observed data is (y_r, A_r^{obs}, z_r), consisting of y_r the vector of outcome data, A_r^{obs} the observed part of the graph, and z_r the vector of network formation covariates. The missing data for each network is A_r^{mis} and recall $G_r = (A_r^{obs}, A_r^{mis})$.

Every network is thought of as a realization of a random network formation process, drawn from a distribution which depends on z_r and parameter $\theta_{0r} \in \Theta_r$. To estimate β_0 we use an argument based on conditional expectations. If θ_{0r} were known for all r, we could estimate a conditional expectation of $w_{ir}(G_r)$ given the observed data,

$$\mathcal{E}_{ir}(A_r^{obs}, z_r; \theta_{0r}) := \mathrm{E}\left[w_{ir}(G_r) | A_r^{obs}, z_r; \theta_{0r}\right].$$

By the properties of conditional expectation, using $\mathcal{E}_{ir}$ in the regression instead of w_{ir} yields consistent estimation of β_0. The least squares estimator is given by[32]

$$\widehat{\beta}_{\text{ols}} = \left(\sum_{r=1}^{R}\sum_{i=1}^{n} \mathcal{E}_{ir}(\widehat{\theta}_r)\mathcal{E}_{ir}(\widehat{\theta}_r)'\right)^{-1} \cdot \sum_{r=1}^{R}\sum_{i=1}^{n} \mathcal{E}_{ir}(\widehat{\theta}_r) y_{ir}.$$

[29] In what follows it is not necessary for y_{ir} to be observed for every node, but it simplifies notation.

[30] E.g., $z_r = \{z_{ir} : i = 1, ..., n\}$ or $z_r = \{z_{ij,r} : i, j \in V\}$ where z_{ir} or $z_{ij,r}$ are covariates for nodes or pairs.

[31] Researchers can either collect simple covariate data from all nodes or from representatives who carry information.

[32] For notational simplicity, assume the regressors are demeaned.

Notice $\widehat{\beta}_{\text{ols}} = \widehat{\beta}_{\text{ols}}(\widehat{\theta}_1, ..., \widehat{\theta}_R)$ depends on $\widehat{\theta}_r$ for all r. A similar but more involved result is true for GMM. We must use the fact that the outcome variable provides information about the likelihood of missing links; this extra complication disappears under the linearity of OLS.

To control the estimation of $\widehat{\theta}_r$, we need to argue not only that it is consistent for θ_{0r}, but uniformly so. That is, $\sup_r \|\widehat{\theta}_r - \theta_{0r}\| = O_{\mathrm{P}}(a_R^{-1} \cdot R_n^{1/b})$, where a_R is the rate of convergence of $\widehat{\theta}_r$ to θ_{0r} for every r, and $b > 1$ is the number of moments that the network formation model has. This imposes a rate requirement on the problem which says that the network-formation parameter needs to be estimated fast enough: $\sqrt{nR} \cdot a_R^{-1} \cdot R^{1/b} \to 0$.

The consistency of $\widehat{\theta}_r$ follows from assumptions on the model of graph formation and the sampling procedure. With missing-at-random data, under assumptions on the graph model, a consistent estimator exists. Consider a model where an edge forms independently, conditional on covariates,

$$\mathrm{P}(A_{ijr} = 1 | z_r; \theta_{0r}) = \Lambda\left(f(z_{ir}, z_{jr})' \theta_{0r}\right),$$

where $\Lambda(\cdot)$ is some link function (e.g., logistic or normal), z_i is a vector of covariates for vertex i, and f is a vector-valued function. For instance, f may be the difference between characteristics of two nodes $f(z_i, z_j) = \|z_i - z_j\|$. If the sampling procedure is orthogonal to the network formation, a random subset of the $\binom{n}{2}$ pairs of nodes is observed. Therefore, $\widehat{\theta}_r$ is consistent.

This model converges with $a_R = n$, since we have on the order of $n(n-1)/2$ observations. The requirement becomes $n^{-1/2} R^{1/2+1/b} \to 0$, so the number of networks must grow sufficiently slower than the number of nodes. In other models, the rate a_R may be different (e.g., $n/\log n$, n^τ for $\tau \in [1/2, 2)$, $\sqrt{n/\log n}$). If the rate is too slow, the requirement for node-level regression may not be met, though usually the requirement for graph-level regressions will be satisfied.

4.2. **Formal Theory for $\widehat{\beta}$.** We begin by establishing that $\widehat{\beta}$ is consistent and asymptotically normal. The main theorem is stated in section 4.2.1, under regularity conditions, including simple high level assumptions about the behavior of $\widehat{\theta}_r$, which we will verify in section 4.4. In Appendix A.2 we discuss the regularity conditions in depth.

We have already introduced the regression environment. We consider the GMM environment of (2.1). Relative to regression, in GMM the value of y affects the conditional expectation of w. Observe (2.1) implies an unconditional moment restriction holds:

$$0 = \mathrm{E}m(X; \beta_0) = \sum_{G \in \mathcal{G}_n} \mathrm{E}\left[m(X; \beta_0) | G\right] \mathrm{P}_{\theta_0}(G)$$

where $X = (y, w(G))$. Let x_r denote the triple of observed data, $x_r := (y_r, A_r^{obs}, z_r)$. By iterated expectations, the conditional function

$$\mathcal{E}_{ir}(x_r; \beta_0, \theta_0) := \mathrm{E}\left[m(X_{ir}; \beta_0) | x_r; \beta_0, \theta_0\right] \tag{4.1}$$

satisfies $\mathrm{E}\mathcal{E}_{ir}(x_r;\beta_0,\theta_0)=0$. Given an observed data series $\{(X_{ir},z_{ir}):i=1,...,n,\ r=1,...,R\}$ and an estimator $\widehat{\theta}_r$ of θ_{0r}, the estimator is

$$\widehat{\beta}_{\mathrm{gmm}} := \operatorname*{argmin}_{\beta\in\mathcal{B}} \left(\mathbb{E}_{n,R}\mathcal{E}_{ir}(x_r;\beta,\widehat{\theta}_r)\right)' \widehat{W} \left(\mathbb{E}_{n,R}\mathcal{E}_{ir}(x_r;\beta,\widehat{\theta}_r)\right)$$

where $\widehat{W}$ is a consistent estimator of W.[33]

In order to compute the conditional moment in (4.1) we need to be able to integrate with respect to a conditional probability for every graph in our sample, $\mathrm{P}_{\beta_0,\theta_0}(A_r^{mis}|x_r)$. Computing the expectation requires a reweighting term,

$$\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r}) = \sum_{A_r^m} m(X_{ir};\beta_0)\mathrm{P}_{\beta_0,\theta_{0r}}(A_r^{mis}|x_r),$$

with $\mathrm{P}_{\beta_0,\theta_{0r}}(A_r^{mis}|x_r) \propto f_{\beta_0}(y_r|G_r)\mathrm{P}_{\theta_{0r}}(A_r^{mis}|A_r^{obs},z_r)$. To be able to utilize this approach, the researcher must make assumptions on the distribution of y given G.

4.2.1. *Asymptotic Distribution.* In this section we show that $\widehat{\beta}_{\mathrm{ols}}$ and $\widehat{\beta}_{\mathrm{gmm}}$ are consistent and asymptotically normally distributed. We define covariance matrices which characterize the asymptotic variance. For linear regression,

$$H_{\mathrm{ols}} := \lim_{R\to\infty} \mathbb{E}_{nR}\left[\mathrm{E}\mathcal{E}_{ir}\mathcal{E}_{ir}'\right] \text{ and } V_{\mathrm{ols}} := \lim_{R\to\infty} \mathbb{E}_R\left[\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\left[\mathcal{E}_{ir}\epsilon_{ir} + \mathcal{E}_{ir}(w_{ir}-\mathcal{E}_{ir})'\beta_0\right]\right)\right],$$

and for GMM,

$$M := \lim_{R\to\infty} \mathbb{E}_{n,R}\left[\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right],\ \Omega := \lim_{R\to\infty}\mathbb{E}_R\left[\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right],$$
$$H_{\mathrm{gmm}} := M'WM \text{ and } V_{\mathrm{gmm}} := M'W\Omega W'M.$$

Theorem 4.1 (Asymptotic Distribution). *Under Assumption A.3,*

(1) *Assumption A.4 implies* $\sqrt{nR}(\widehat{\beta}_{\mathrm{ols}}-\beta_0) \rightsquigarrow \mathcal{N}\left(0, H_{\mathrm{ols}}^{-1}V_{\mathrm{ols}}H_{\mathrm{ols}}^{-1}\right)$.
(2) *Assumption A.5 implies* $\sqrt{nR}(\widehat{\beta}_{\mathrm{gmm}}-\beta_0) \rightsquigarrow \mathcal{N}\left(0, H_{\mathrm{gmm}}^{-1}V_{\mathrm{gmm}}H_{\mathrm{gmm}}^{-1}\right)$.

Intuitively, if we can uniformly replace $\widehat{\theta}_r$ with θ_{0r}, since conditional expectations are centered correctly and, under regularity conditions, also satisfy central limit theorems if the unconditioned random variables do, the estimator is consistent and normal. While we wrote the theorem for vertex-level analysis, similar results with modified regularity conditions extend to regressions at the graph-level, edge-level, vertex-triples, etc. Each will allow for different amounts of interdependency in the graph formation process. To be concrete, under the above normalizing assumptions, graph-level regression converges at $\sqrt{R}$ while edge-level regression converges at $\sqrt{\binom{n}{2}R} = n\sqrt{R}$.

To build further intuition, we comment on what could go wrong. First, for GMM, if one estimates the conditional expectation without reweighting, unless the model was additively separable, $\widehat{\beta}_{\mathrm{gmm}}$ would be inconsistent. Second, there are several reasons why uniform estimation may fail: the size of the networks relative to the number of networks may be too

[33] In the case of maximum likelihood where $\mathcal{E}$ is the conditional score, $W=I$.

small, the network formation process may have $\dim(\Theta_r)$ exploding too fast, and the level of interdependency in the random graph processes may be too high. We provide a more detailed discussion in section 4.5.

4.3. **Estimation in Practice.** We describe the estimation algorithm for linear regression.

Algorithm (Estimation of $\widehat{\beta}_{\text{ols}}$).

(1) Use (z_r, A_r^{obs}) to estimate $\widehat{\theta}_r$ based on the assumed network formation model.
(2) Estimate $\mathcal{E}_{ir}(A_r^{obs}, z_r; \theta_{0r}) := \mathrm{E}\left[w_{ir}(G_r)|A_r^{obs}, z_r; \theta_{0r}\right]$.
    (a) Given (z_r, A_r^{obs}), for simulations $s = 1, ..., S$, draw $A_{r,s}^{mis*}$ from $\mathrm{P}_{\widehat{\theta}_r}(A_r^{mis}|A_r^o, z_r)$.
    (b) Construct $w_{ir}(G_{rs}^*)$ where $G_{rs}^* = (A_r^{obs}, A_{r,s}^{mis*})$.
    (c) Estimate $\widehat{\mathcal{E}}_{ir}(A_r^{obs}, z_r; \widehat{\theta}_r) := \frac{1}{S}\sum_{s=1}^S w_{ir}(G_{rs}^*)$.
(3) Estimate $\widehat{\beta}_{\text{ols}}$ using data $\{(y_{ir}, \widehat{\mathcal{E}}_{ir}(A_r^{obs}, z_r; \widehat{\theta}_r)) : i = 1, ..., n,\ r = 1, ..., R\}$.

The GMM algorithm is similar, requiring a reweighting term. We provide an overview of standard errors and estimation methods in Appendix G, though a theoretical development of them is well-beyond the scope of this paper. In practice, clustering at the graph level in vertex-level regressions and using heteroskedasticity robust standard errors for network-level regressions perform well, though we have explored various bootstrapping procedures (available upon request).

4.4. **Formal Theory for $\widehat{\theta}_r$.** In this section we discuss the uniform estimation of the network formation model parameters. We are interested in the joint convergence of $\sup_r \left\|\widehat{\theta}_r - \theta_{0r}\right\|$ in the sense of Assumption A.3.3. The literature on consistently estimable network formation models is young and limited, though has seen considerable growth in econometrics since the first version of this paper. Many models of network formation lack asymptotic frames (see, e.g., exponential random graphs models (ERGMs)) that allow for consistently estimable parameters, or the models may not even be projective (Shalizi and Rinaldo, 2012). So larger networks do not lead to tighter parameter estimates and seeing a slice of the network may not allow recovery of the true parameter. There are a few classes of models known to be consistent, and we discuss several as examples below. Given how nascent this literature is, it is useful to reflect on a simple, checkable sufficient conditions for joint convergence so that one could check new models as they develop. After this, we discuss three common classes of network formation models and check the condition that can be used in graphical reconstruction. The examples have been chosen to provide intuition about different problems that may arise.

We have a collection of network formation models which maximize criterion functions, $\theta_{0r} = \arg\max_\theta Q_{(r)}(\theta_{0r})$. We estimate these parameters with a collection of empirical criterion functions, $\widehat{Q}_{(r)}(\theta_r)$, with $\widehat{\theta}_r = \arg\max_\theta \widehat{Q}_{(r)}(\theta_r)$. The lemma following is analogous to Hahn and Newey (2004); we include it here to point the reader to what we need for our procedure to work well.

LEMMA 4.1. *Let* $V_{(r)}(\theta_r) := \nabla_\theta Q_{(r)}(\theta_r)$ *and* $\widehat{V}_{(r)}(\theta_r) := \nabla_\theta \widehat{Q}_{(r)}(\theta_r)$. *Assume the following.*

(1) $\forall r$, $Q_{(r)}(\theta_r)$ *has unique maximum* θ_{0r}; Θ_r *is compact;* $Q_{(r)}(\theta) \in C^2(\Theta)$; $\sup_\theta |\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)| = o_{\mathrm{P}}(1)$.
(2) *The criterion functions uniformly converge in the sense that for some* $v > 0$
$$\mathrm{P}\left(\sup_{r \le R} \sup_{\theta \in \Theta_r} \left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \ge \eta\right) = o(a_R^{-v}).$$
(3) *There exists a sequence of constants* (a_R) *such that (i) for all* r, $a_R \cdot \widehat{V}_{(r)}(\theta_{0r}) = O_{\mathrm{P}}(1)$; *(ii) for some* $b > 1$, $\sup_{r \le R} \mathrm{E}\left\|a_R \cdot \widehat{V}_{(r)}(\theta_{0r})\right\|^b < \infty$.
(4) $\nabla \widehat{V}_{(r)}(\theta_r)$ *satisfies a Lipschitz condition with coefficient* B_r, $\sup_r \|B_r\| = O_{\mathrm{P}}(1)$.
(5) *The Hessian satisfies* $\sup_r \left\|\nabla \widehat{V}_{(r)}(\theta_r) - \nabla V_{(r)}(\theta_r)\right\| = o_{\mathrm{P}}(1)$.

Then $a_R \cdot R^{-1/b} \cdot \sup_{r \le R} \left\|\widehat{\theta}_r - \theta_{0r}\right\| = O_{\mathrm{P}}(1)$.

This provides checkable conditions to ensure Assumption A.3 holds and therefore the main theorem holds. The argument comes from a usual first order expansion argument. Condition 1 adds extra smoothness to a standard assumption for consistency. Condition 2 requires that all the criterion functions $\widehat{Q}_{(r)}(\theta)$ uniformly lie in an η-"sleeve", $[Q_{(r)}(\theta) - \eta, Q_{(r)}(\theta) + \eta]$; in practice this is argued by applying union bounds and controlling interdependencies across summands in the objective function. Condition 3 provides a rate of convergence of the first-order term and a moment requirement. Condition 4 requires an envelope condition for the third derivative of the objective. Condition 5 requires uniform convergence of the Hessian. Below, we check that Lemma 4.1 holds under low-level assumptions.

4.4.1. *Classes of Models.* The goal of this section is to outline several network formation models that could fit into this framework. This is not meant to be exhaustive but instead give the reader a sense of how various models could be used.

Class 1: Conditional Edge Independence Models. We begin by considering a class of models in which edges form independently, given covariates. This is the most common class of model used in the literature (see e.g., Jackson, 2008b, Christakis et al., 2010, Goldsmith-Pinkham and Imbens, 2011, and Santos and Barrett, 2008). Let Ξ be a set consisting of all pairs ij. Ξ is implicitly indexed by n and has $n(n-1)/2$ elements. We denote an element $s \in \Xi$ and, when referencing explicitly which pair it corresponds to, we write $s = s_{ij}$. Let z_s denote a covariate for the pair of nodes s_{ij}. Examples include whether two villagers are of the same caste, the distance between their households, etc. The probability that an edge forms in graph r is

$$\mathrm{P}\left(A_{sr} = 1 | z_{sr}; \theta_{0r}\right) = \Phi\left(z_{sr}' \theta_{0r}\right) \tag{4.2}$$

where $\Phi(\cdot)$ is some link function. This framework allows us to consider undirected graphs, directed graphs, and models in which nodes have to agree for a link to form. The undirected case is clear. If the graph formation model is directed, then Ξ consists of all $n(n-1)$ ordered pairs of ij.[34]

We maximize the log-likelihood, $|\Xi|^{-1}\sum_{s\in\Xi} q(X_s;\theta_r)$, $X_s = [A_s, z_s']$, with summand

$$q(X_s;\theta_r) = A_{sr}\log\Phi(z_{sr}'\theta_r) + (1-A_{sr})\log\left(1-\Phi(z_{sr}'\theta_r)\right).$$

To be able to apply Lemma 4.1, we need to control the interdependence in the covariates z_{sr}. We assume that the set of nodes itself has an embedding into an integer lattice, $\Lambda \subset \mathbb{Z}^d$. Let $t \in \Lambda$ denote a generic element, and when referencing the corresponding node we write $t = t_i$. To build intuition imagine that the nodes are embedded in $\mathbb{Z}^2$ as analogous to geographic placement; households in a village are placed on a grid on the ground and certain households are closer to others. This closeness determines the covariance of their other characteristics. Then, every node is given a random covariate z_i, $t_i \in \Lambda$. The pair-level covariate z_{ij} is given by $z_{ij} = f(z_i, z_j)$ for some function $f(\cdot,\cdot)$. The interdependency in the node-level covariates will translate to interdependencies among the edge-level covariates which is what we will ultimately use in our argument. We will need to assume that the level of interdependence goes to zero as the distance between the two subsets goes to infinity. Figure 4 provides an illustration.

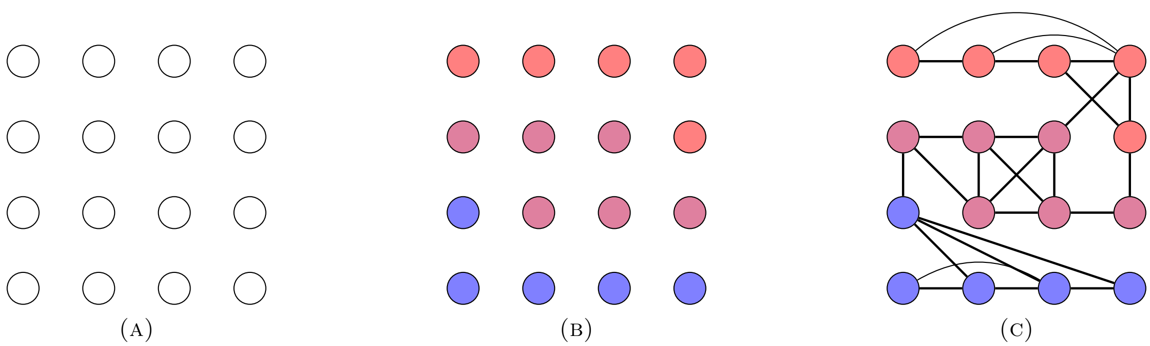


FIGURE 4. This figure presents a schematic of this model. In Panel A, nodes are on a lattice. In Panel B, nodes draw covariates z_i correlated over the lattice, here represented in colors (red, purple, blue). In Panel C, nodes are linked with higher probability if they have closer traits ($\|z_i - z_j\|$ is low).

This assumption we make is analogous to those made in time series and spatial econometrics contexts. We require that the random fields z_r satisfy uniform mixing requirements where, as the distance between the sites of two random variables increase, the level of interdependency decays quickly. The assumption on $f(z_i, z_j)$ is not very restrictive. The most natural example is a covariate based on the difference in characteristics of nodes i and j: $z_{ij} = \|z_i - z_j\|$.

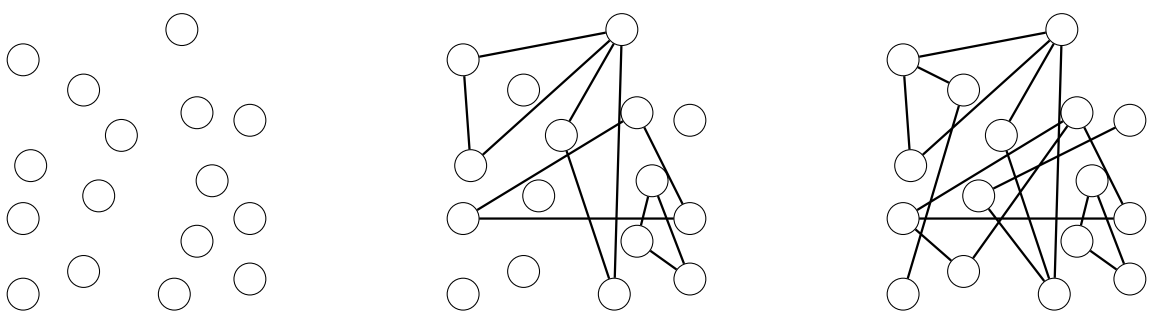
[34]When the model is undirected but both nodes need to agree, one may use a model such as $A_{ijr} = \mathbf{1}\{z_{ij}'\theta_{0r} - \epsilon_{ijr} \geq 0\}\cdot\mathbf{1}\{z_{ji}'\theta_{0r} - \epsilon_{jir} \geq 0\}$ with link function $\Phi\left(z_{sr}'\theta_{0r}\right) := \Psi(z_{ij}'\theta_{0r})\Psi(z_{ji}'\theta_{0r})$, where $\Psi(\cdot)$ is the cdf of ϵ. One may even want to assume ϵ_{ij} and ϵ_{jr} being jointly normal.

Proposition 4.1. *Assumptions A.6 and A.7 imply the conditions of Lemma 4.1.*

Until now we have not discussed the role of random sampling. It is easy to see with random sampling of nodes (G^S or $G^{|S}$) or random sampling of pair data A_{ij} that the criterion function $Q_{(r)}(\theta) := \lim_{n\to\infty} \left|\Xi^S\right|^{-1} \sum_{s\in\Xi} \mathrm{E}[q(X_{sr};\theta)\mathbf{1}\{s \in \Xi^S\}]$ is minimized at true parameter θ_{0r}. For instance, under the star subgraph we have $\mathrm{E}[q(X_{sr};\theta)\mathbf{1}\{s \in \Xi^S\}] = (1-(1-\psi)^2)\mathrm{E}q(X_s;\theta_{0r})$ while $\left|\Xi^S\right| = (1-(1-\psi)^2)\binom{n}{2}$. More generally, if the sampling procedure is known, then in such a model augmenting the likelihood to account for the sampling will produce consistent estimates.

Class 2: Subgraph Generated Models (SUGMs). One limitation with models in which the covariates direct the correlation between links is that there still may no be enough clustering. Chandrasekhar and Jackson (2016) provide another class of models, *subgraph generated models* (SUGMs), in which this is accounted for. Here we describe a special case of one of these models.

Let a network be formed in the following manner. Each $\binom{n}{2}$ pairs of nodes are considered and independently, with probability $p_{0L,r}$, a link forms. Similarly, each $\binom{n}{3}$ triple of nodes are considered and independently, with probability $p_{0T,r}$, a triangle forms. The observed graph G_r is the union of these processes. Figure 5 provides an illustration.



(a) (b) (c)

Figure 5. This figure presents a schematic of the links and triangles SUGM. In Panel A, nodes are placed arbitrarily, as there is no ex ante natural embedding. In Panel B, triads are drawn uniformly at random. In Panel C, links are drawn uniformly at random, and the resulting network is presented.

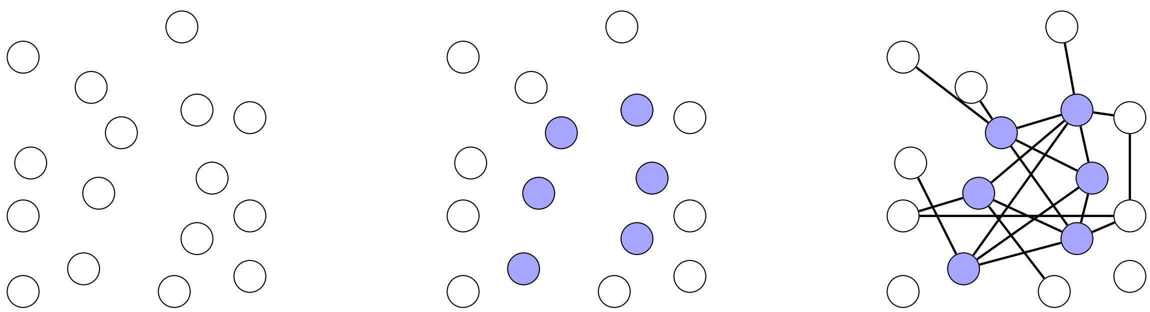
There are a number of microfoundations for such models: mutual consent, directed search for group formation, among others give rise to such models. Practically speaking, these models have the advantage of, in a simple and naturalistic way, coding in correlation in link structure and therefore being able to generate sparse and clustered networks. This is important because empirical data is typically sparse and clustered. Chandrasekhar and Jackson (2016) shows that a links and triangles SUGM outperforms a conditional edge independence model with numerous demographic covariates in terms of matching various network features: clustering, average path length, maximal eigenvalue, etc.

Proposition 4.2. *Assumption A.8 implies the conditions of Lemma 4.1.*

Class 3: Group Models. By allowing for an increasing number of parameters, a network formation model may be able to better and more flexibly describe the random graph process. Models of this vein are discussed in Bickel and Chen (2009), among others, who provide a discussion of what they call a nonparametric view of network models. Our specific example comes from Chatterjee et al. (2010), who study an environment in which the degree distribution is the sufficient statistic for graph formation: given $(d_1, ..., d_n)$, one estimates a formation model.[35] Following Diaconis and Freedman (1984) they show the network is described by

$$\mathrm{P}(A_{ij} = 1) \propto \exp\left(\theta_{0i} + \theta_{0j}\right),$$

which is a model that allows the number of parameters to grow at a $\Theta(n)$ rate. Obviously we cannot have a parameter per node because we will not have observations for a number of nodes, especially under induced subgraph sampling.



(A) (B) (C)

Figure 6. This figure presents a schematic of the groups model. In Panel A, nodes are placed arbitrarily, as there is no ex ante natural embedding. In Panel B, we denote $\theta_{0,blue}$ in blue: these nodes have a high fixed-effect value, Meanwhile $\theta_{0,white}$ nodes have a low fixed-effect value. In Panel C, links are drawn independently, with probability proportional to the sum of the fixed effects and the resulting network is presented. There are many links between the blue nodes, fewer links between the blue and white nodes, and rarely any links between white nodes.

So, we tweak this framework to our environment and assume there are k_n categories of nodes. For instance, if graph formation depends on two characteristics, gender (male/female) and education (high/low), there are four such categories. By allowing k_n to grow rapidly

[35]Conditional on the degree distribution there is no information about the model from the actual network data.

with n, we can capture substantial variation in the formation of the network. We allow $k_n = \Theta(n)$. Define an equivalence class of nodes: if i and j are in the same class, then they have the same parameter, $\theta_{0i} = \theta_{0j}$. In our example, two individuals in the same category (e.g., female and high education) are governed by the same parameter $\theta_{0,female,high}$. If we have q_n characteristics with uniformly bounded values (e.g., two genders, a bounded number of education levels), the number of categories can grow at $q_n = \Theta(\log n)$, which yields $k_n = \Theta(n)$. One can think of this model as having group fixed effects with a growing number of groups. It turns out that with probability approaching one, $\sup_{r \le R} \left\| \widehat{\theta}_r - \theta_{0r} \right\|_\infty \lesssim \sqrt{\log n/n}$ which is a very slow rate, though expected given how rapidly we are increasing the parameter dimension. Figure 6 provides an illustration.

Proposition 4.3. *Let the maximum coordinate value of $\theta_r \in \Theta_r$ be uniformly bounded over all r, $R = o(n \cdot \log^{-1} n)$, and $k_n = \Theta(n)$, $k_n < n$. Then, under stratified random sampling with either the induced or star subgraph and Assumption A.4 or A.5, the conclusion of Theorem 4.1 holds.*

This example shows that even when we are adding parameters at rate n, graphical reconstruction is possible in network-level analyses. Here $a_R = \sqrt{n/\log n}$ and therefore $a_R^{-1} \cdot \sqrt{R^{1+2/b}} \to 0$. Meanwhile, the sufficient condition is not met for vertex-level analysis as $a_R^{-1} \cdot \sqrt{nR^{1+2/b}} \to \infty$. This example provides an illustration of both the strengths and limitations of graphical reconstruction by testing the limits as we add dimensions at the same rate as the number of nodes.

4.5. **Discussion.**

In this section we have developed a general method to consistently estimate the economic parameter using graphical reconstruction. The method allows the researcher to estimate network effects using a general set of network statistics, such as eigenvector centrality, where no analytical solutions are available.

Of course, we may be interested in how misspecification of the network formation model affects graphical reconstruction. In practice, we do not know the family of models which generated the empirical networks. Clearly, misspecification is problematic only to the extent of the covariance between the conditional expectation of the misspecified model and its deviation from the true model. While this is not easy to analytically characterize, it does suggest that the model one needs relates to the network statistic one is interested in studying. For instance, graphical reconstruction with Erdos-Renyi style models may be sufficient to study questions pertaining to the degree distribution, but may perform poorly if one is interested in clustering. Numerical simulations confirm precisely this intuition, suggesting that chosen models ought to be a function of the statistic of interest.

Another natural avenue for future work could assess the trade-off of dense versus sparse network formation models. Loosely speaking, in dense models, there is more information but on the other hand, identification issues can be thornier.

5. Numerical Experiments

This section reports the results of numerical simulations that characterize the biases due to sampling as well as the behavior of the analytical and graph reconstruction estimators.

5.1. **Simulation Setup.** We specify a data-generating process for a set of random graphs and outcome data, and then carrying out the following steps.

Algorithm (Simulation).

(1) Generation of data.
   (a) Draw R networks from the network formation families (below).
   (b) Generate outcome data from a model with β_0 and data-generating process $(y, \epsilon)|G; \beta_0$.
   (c) For each graph G_r construct sampled graphs $\{G^S_{rb}, G^{|S}_{rb} : b = 1, ..., B\}$.
(2) Estimation of $\widehat{\beta}$ using $\{G^S_{rb}, G^{|S}_{rb} : b = 1, ..., B\}$.
   (a) Estimate $\widehat{\beta}_b(G^{|S})$ and $\widehat{\beta}_b(G^S)$.
   (b) If applicable, estimate the adjusted estimator $\widetilde{\beta}_b(G^{|S})$ and $\widetilde{\beta}_b(G^S)$.
   (c) Estimate the graphical reconstruction estimators.
(3) Perform (1)-(2) for $\psi \in \{1/4, 1/3, 1/2, 2/3\}$.

We generate networks of $n = 250$ nodes using the following simple conditional edge independence model. We set parameters such that the average degree, clustering, path length, maximal eigenvalue and variance of the eigenvector centrality distribution from networks in our simulations mimics those moments in the empirical Indian networks data-set. Dividing the set of nodes into 6 approximately equally sized groups, we place those groups on a line, indexed from 1 to 6. The probability that an edge formed between two members within the same group is high. The probability that an edge formed between two members of two different groups declines in the cross-group distance, represented by the difference in the indexed location of those groups on the line. Formally, let $g(i)$ denote the group of vertex i. We set

$$\mathrm{P}(A_{ijr}|z_{ijr}) = z'_{ijr}\theta_{0r}.$$

θ_{0r} is a $\binom{6}{2}$-vector with elements $\theta_{0r,lm}$ with $1 \le l < m \le 6$ and z_{ijr} is a $\binom{6}{2}$-vector with $z_{ijr,lm}$ with $1 \le l < m \le 6$. $\theta_{0r,lm}$ is the probability that a member of group l is linked to a member of group m. The lm-component of z_{ijr} is a dummy for whether i and j are in groups l and m respectively, $z_{ijr,lm} = \mathbf{1}\{g(i) = l\}\mathbf{1}\{g(j) = m\}$. In order to generate $\theta_{0r,lm}$ we use a simple distance function, with $\theta_{0r,lm} = p_r(1 - |l - m|/6)$ where p_r is a uniform random variable chosen such that the average degree generated mimics the average degree from the empirical application.

We have conducted simulations for alternative formation models, such as one in which covariates are generated by an autoregressive process and the edge formation probability is governed by a logistic regression. Another example is the subgraph generated models (Chandrasekhar and Jackson, 2016). Results are qualitatively and quantitatively similar.

5.2. **Regression of Outcomes on Network Characteristics.** We simulate and estimate a model with heteroskedastic residuals, $y_{ir} = \alpha_0 + w_{ir}(G_r)\beta_0 + \sigma_0 \cdot u_{ir}$, where $(\alpha_0, \beta_0) = (1, 2)$,

$$u_{ir} := \mathcal{N}\left(0, \sigma_{ir}/\sqrt{\hat{\mu}^2_{\sigma_{ir}}}\right), \ \sigma_{ir} := 3\frac{w_{ir} - w_{ir\,\min}}{w_{ir\,\max} - w_{ir\,\min}} + 0.2, \text{ and } \hat{\mu}^2_{\sigma_{ir}} := \mathbb{E}_{n,R}\left[\sigma^2_{ir}\right].$$

This formulation creates a fan-like heteroskedasticity. We then can easily set the R^2 of the regression to approximately 0.3 by defining $\sigma_0^2 := (1/\mathrm{R}^2 - 1) \cdot \mathbb{E}_{n,R}(\tilde{y}_{ir} - \bar{\tilde{y}}_{ir})^2$ for $\hat{y}_{ir} = \alpha_0 + w_{ir}(G_r)\beta_0$.

Columns 1-5 of Tables 7 and 8 show the estimation bias, in percentages, for regression parameters when using sampled network data for a variety of network statistic regressors. Table 7 shows the biases when estimating regressions at the network level while Table 8 shows the biases when estimating regressions at the node level.

At the network level, we consider average degree, graph clustering, graph span, average path length, and λ_1. In addition, we show results for the standard deviation of the eigenvector centrality distribution and the spectral gap. The eigenvector centrality represents how important a node is in information transmission (Jackson, 2008b) and the spectral gap of a graph characterizes how rapidly diffusion processes on networks spread (Chung, 1997). The latter is closely related to the expansiveness of a network that Ambrus et al. (2010) show characterizes good risk-sharing properties.

At the node level, we show results for the degree, clustering coefficient, and eigenvector centrality of a node. Moreover we consider two regressions which characterize how far a node i is from another node j. We select a random node j (corresponding to a randomly treated node in an experimental setting) and generate a regressor which is the path length from i to j. In addition, we partition the nodes into two subsets which communicate the most within themselves and least across the sets. We say i is on the same side of the spectral partition of j if they are in the same subset. This partition is related to the spectral gap (Chung, 1997) and therefore has implications for the Ambrus et al. (2010) approach to characterizing risk-sharing.

Overall we find that sampling the network leads to significant biases. Figures 7 and 8 present the corresponding results, graphically. Here we discuss the biases at 1/3 sampling for the graph and node level. At the graph level the maximum bias is 260% (λ_1, induced subgraph), the mean is 90.9%, and the minimum is 15% (clustering, induced subgraph).[36] The biases include expansion bias in the cases of degree, maximal eigenvalue, spectral gap, and graph clustering (for the star subgraph). The node-level regressions exhibit a similar

[36] When we looking at maximum, mean, and minimum, we are interested in the magnitude of the biases, so our discussion focuses on the absolute value of the bias.

pattern: the maximum bias magnitude is 91% (degree, induced subgraph), the mean is 63%, and the minimum is 7% (same side of the spectral partition).

In Figure 7, we also present the results using the analytical corrections from section 3.[37] We find that the adjusted regression estimators perform uniformly better than the unadjusted estimators. The biases are usually low. Overall, the analytical corrections improve 100% of the parameter estimates in the simulations. At 1/3 sampling, the mean reduction in bias percentage when comparing the analytical correction to the raw network statistic is 69pp with a median reduction of 69pp and a maximum of 243pp. Similarly Figure 8 applies the analytical corrections which were derived for the case of graph level regression, to node-level analysis; the results are not motivated by theory and of course are mixed.[38]

We also consider the graphical reconstruction estimators in Figures 7 and 8. It nearly uniformly outperforms the estimator using the sampled data alone. Biases are mostly very low across a number of linear and nonlinear network statistics. For illustration we discuss examples with 1/3 sampling: at the graph level the median bias is 5.7%, the minimum is 0.6%, and the reconstruction estimator reduces the bias in 54 of the 56 parameters estimated in Figure 7. The mean reduction in bias is 73pp and the maximum reduction is 254pp. We find similar results at the node level. The median bias is 1.4% and graph reconstruction reduces the bias in 100% of the parameters estimated in the table. Furthermore, the median reduction in bias is 62pp with a maximum of 85pp. Not surprisingly, at a given sampling rate reconstruction with G^S performs uniformly better than with $G^{|S}$. The effective share of edges observed in the star subgraph is $1-(1-\psi)^2$ but is ψ^2 in the induced subgraph. Typically 2/3 sampling with $G^{|S}$ (4/9 share of the edges observed) yields a reconstruction procedure which is only as good as 1/4 sampling with G^S (7/16 share of the edges observed).

In Table 1 we study the behavior of significance testing and provide evidence that graphical reconstruction may often increases t-statistics.[39] Specifically, we present the ratio of the t-statistic under graphical reconstruction to the t-statistic under the naive estimator using the sampled network statistic. We find that at the network level across 86% of the cases the t-statistic increases (48 of 56 estimated parameters) and at the node level across 96% of the cases (46 of the 48 estimated parameters) graphical reconstruction yields a higher t-statistic than the naive estimator. Moreover, we find that the average ratio of the t-statistic of reconstruction to the naive estimator is high.

5.3. **Regression of Outcomes on Network Neighbors' Outcomes.** Table 2 presents the results for simulations for the model of equation (2.3) with $(\alpha_0, \rho_0, \gamma_0, \delta_0) = (1, 0.5, 2, 0.5)$. We use three specifications to demonstrate the emergence of biases in peer effects regression due to two distinct causes: correlation of the instrument with the errors-in-variables problem

[37]In Table H.1 of Appendix H we show an example of an analytical correction that involves estimating $\widehat{\sigma}_v^2$.
[38]The correction working for degree with G^S is mechanical since there is no mismeasurement for d_i with $i \in S$.
[39]Note that this is a numerical result and not a theoretical one. The results may be specific to network formation models and statistics examined.

and a weak instrument/finite sample problem induced by sampling. The table presents the mean bias percentage as well as the standard error of the bias.

We present three methods of estimating peer effects with sampled data and one correction. First, we show an estimate of the peer effect model with the network given by the induced subgraph. Second, we present an estimation where (y, x) are known for every node, but the network used is the star subgraph. Finally, we present the same specification but only allow the researcher to have covariates for surveyed nodes. Each specification exhibits biases.

We vary the number of networks and disturbance size across three models to study how the bias varies. In Model 1, we use one network with 250 nodes per simulation drawn from the aforementioned model and set the number of simulations to $50,000$. In Model 2, we use one network with 250 nodes per simulation and use $50,000$ simulations, but reduce the variance of the disturbance. Model 3 presents results a from $2,500$ simulations of 20 networks each with 250 nodes, drawn from the model.[40]

All specifications show significant bias in the estimates of the endogenous and exogenous peer effects. Comparing Panels A and B of the first and second set of columns shows the biases are greater when there is more noise in the system. Moreover, comparing Panels A and B of columns 1-5 and 11-15 shows that increasing the number of graphs in the estimation from 1 to 20 only modestly reduces the bias due to sampling. Non-trivial biases which remain.

Overall, the analytical correction performs well. In the Models 2 and 3, the estimates are essentially unbiased across all sampling levels presented. Moreover, the analytical correction for Model 1 exhibits negligible bias for sampling rates of 2/3 and 1/2. However, biases emerge at very low sampling rates, 1/3 and 1/4, in the case of Model 1. Furthermore, as evidenced by the standard errors at 1/4, the estimates are extremely unstable.

To measure whether there is a weak instruments problem, in Panel C we display a generalization of the concentration parameter of the first stage, allowing for interdependence in the variance following Kleibergen (2007).[41] The intuition is that in these networks, even for the analytical correction there is measurement error in the instrument. Since the number of connections to neighbors and second neighbors in a network is low, the amount of noise in the first stage increases.[42] Panel C shows that the concentration parameter is very low for the first stage estimates in Model 1, especially at low sampling levels. Moreover, once the number nodes in the network is high enough or the amount of independent data (20 networks) is high enough, the concentration parameter is extremely high. In these cases our analytical correction removes the bias entirely while biases remain with the sampled estimators.

[40]The number of simulations was chosen to roughly equate the computation time, on the order of $n^4 \cdot \#$ of simulations, for each of the three specifications.
[41]For a first stage $X = Z\pi + v$, we use the generalized concentration parameter $\mu^2 := \pi' \Sigma_\pi^{-1} \pi$ where $\widehat{\pi} = (Z'Z)^{-1} Z'X$ and $\Sigma_\pi = \mathrm{var}(\widehat{\pi})$.
[42]The extent to which this matters can be seen by noticing that the concentration parameter is 2 for $\psi = 1/4$, while if the signal to noise ratio had stayed the same in the first stage, the concentration parameter should have only decreased from 16 to 4.

5.4. **A Model of Diffusion.** We numerically study the Jackson and Rogers (2007b) model of diffusion and present the results in Table 3. In Panel A we use the aforementioned simulated network data to generate a model with $\beta_0 = e^{-2}$.[43] Columns 1-5 present evidence of severe expansion bias in the estimates $\widehat{\beta}$ when using sampled data. At 1/3 sampling, the transmission parameter is overestimated by 250% when we study the induced subgraph and 85% when we turn to the star subgraph. Columns 6-10 presents the graphical reconstruction results; the procedure removes the entire bias.

5.5. **Robustness to Misspecification.** To investigate how well the procedure works with empirical data, where we do not know the data generating processes, we conduct numerical experiments using the networks of the Banerjee et al. (2013) data-set, described in greater detail in section 6. We repeat the simulation algorithm of section 5.1 with the only difference coming in step 1(a). Instead of generating networks from the aforementioned model, we take 50 independent draws with replacement from the Banerjee et al. (2013) data-set.[44] When we fit a network formation model in step 2(c), we use the model given by (4.2). We use as covariates the GPS distance between households as well as the difference in the number of rooms, beds, roofing material type, and electricity access. Table 4 presents summary statistics from graphical reconstruction exercises analogous to those of Figures 7 and 8. We find that graphical reconstruction reduces the bias in 98% of the network statistics when using the induced subgraph and 100% when using the star subgraph. In addition, the median bias is 9% with the star subgraph when using reconstruction with a median reduction of bias of 23pp. Similarly, the median bias is 32% with the induced subgraph and the median reduction in bias is 32pp. We have also conducted exercises (available upon request) where we do graphical reconstruction but force all villages to be drawn from a model with common parameter θ_0 instead of $\theta_{01}, ..., \theta_{0R}$: there is little reduction in bias, showing that allowing for heterogeneity is essential.

Panel B of Table 3 presents the results from numerical experiments done for the Jackson and Rogers (2007b) model of diffusion using the empirical networks instead of simulated networks. We find that at 1/3 sampling graphical reconstruction yields biases of 5% and 8% for the star and induced subgraphs, respectively. Taken together, the results of these exercises suggest that even when allowing for network formation model misspecification, graphical reconstruction typically outperforms what the researcher otherwise would have estimated.

[43]This choice was motivated by Jackson and Rogers (2007b) who numerically show this corresponds to a 20% steady-state rate of diffusion. This matches the microfinance take-up rate in our empirical application.

[44]We treat the networks as if they are fully-observed. In step 1(c) of the algorithm we sample each graph at rate ψ. The authors of Banerjee et al. (2013) currently are obtaining a 100% network sample in a resurvey.

6. Empirical Application

This section presents an empirical application using data from Banerjee, Chandrasekhar, Duflo, and Jackson (2013), which studies how households' decisions to participate in microfinance diffuses through village networks. We use detailed demographic and social network data in 43 villages in Karnataka, India, which range from a 1.5 to 3 hour's drive from Bengaluru. The data was collected six months before a microfinance institution started its operation in those villages. The networks are randomly sampled at $\sim 46\%$.

The key outcome variable is the microfinance take-up decisions of households in the network. Information about microfinance access is typically spread by members and the MFI has administrative data which allows us to observe the diffusion of membership. Identification is based on the principle that the MFI followed the same procedure in informing villagers about microfinance in each village. The MFI identified a collection of pre-set individuals in the village (*anganwadi* teachers, shop keepers, etc.), informed them about the program in a private meeting, and then asked them to invite individuals to an information session. Banerjee et al. (2013) contend that this scheme provides arguably exogenous variation in the centrality of those households.

To account for the partial sampling, we assume that an edge forms between a pair of households conditionally independently, given a set of covariates (GPS coordinate Euclidean distance between the two households and the difference in the number of beds, number of rooms, electricity access, and roofing material of the two households). We estimate the model separately on each village using a logistic regression in which the observed data between two households are coded as 1 (connected) and 0 (not connected). We have also repeated the exercise but using a more sophisticated model, the subgraph generated model (Chandrasekhar and Jackson, 2016). Results are qualitatively similar.

Panel A of Table 5 reports estimates of village-level regressions where the microfinance take-up rate in a village is regressed on network characteristics. Columns 1-4 presents regressions of microfinance take-up on network statistics, suggested by diffusion theory to be associated with take-up. Column 1 shows the regression of take-up rate on the average eigenvector centrality of the set of initially informed households. Diffusion theory suggests that eventual take-up of microfinance ought to be higher when the first people to be informed are more central. The increase of the average centrality in the set of nodes by 0.1 corresponds to a 16.3pp increase in take-up rate when using the sampled data; graph reconstruction places this estimate as a 24.2pp increase in take-up rate. If the initially informed households were from the 75th percentile of the centrality distribution as compared to the 25th percentile, this represents a 7.5pp increase in microfinance take-up when estimated using reconstruction as compared to a 4.5pp increase when using the sampled data. Recalling that the average take-up rate is 18.49%, this suggests that sampling the network causes significant underestimation of the network effect. Column 2 presents the regression of take-up on the average path length. If it takes one extra step on average to traverse the graph, this corresponds to a 5.4pp decrease in take-up of microfinance, according to the sampled network, though reconstruction suggests that the estimate ought to be a 9.3pp decrease (with a t-statistic of 1.56). Furthermore, consistent with the results of Table 1, the t-statistics associated with the estimates typically increase after reconstruction, suggesting that the researcher can better detect anticipated effects with reduced measurement error.

In Panel B, we turn to household-level regressions where whether a household joins microfinance is regressed on network characteristics. Column 1 reports a regression on the eigenvector centrality of a node (more central nodes should be more likely to hear about the opportunity, under a simple diffusion model). Graphical reconstruction only yields a modest change in this example: an increase in 0.1 of a household's eigenvector centrality corresponds to a 5.5pp increase in take-up likelihood using the sampled data and a 6.6pp increase in take-up likelihood using graphical reconstruction. Column 2 provides a more stark example in a regression of take-up on the inverse social distance of a household to the set of initially informed households. The network effect more than doubles when using graphical reconstruction. Being distance 1 versus 4 increases the probability of joining microfinance by 3.4pp under graphical reconstruction but only 1.6pp using the star graph. In addition, the point estimate is not statistically significant at conventional levels (with a t-statistic of 0.9), but graphical reconstruction establishes that zero is nearly excluded from a 90% confidence interval (with a t-statistic of 1.62).

Finally, in Panel C we consider the regression of a household's decision to join microfinance or not on the sum of its neighbors' decisions.[45] The of ρ estimate corresponds to the impact of one neighbor joining microfinance on the probability that a household joins microfinance. Column 2 displays the parameter estimate of the effect of the exogenous covariate, γ, whether a household is initially informed about microfinance. Column 3 displays the parameter estimates of exogenous network effect, δ. This estimate describes the impact of one extra neighbor being a initially informed on a household's likelihood of joining microfinance. We focus on column 1 as the endogenous network effect is the key parameter of interest. The star subgraph data suggests that a one neighbor's take-up corresponds to a 2.7pp decrease in the likelihood of a household joining microfinance (which a researcher might interpret as a substitution effect). Meanwhile, the star subgraph data where we use the microfinance data and injection point data only for sampled households suggests that a one more neighbor's take-up corresponds to a 4.7pp increase in the likelihood of a household taking up (which a researcher might interpret as an information or endorsement effect). Finally, the analytical correction shows that a one neighbor's take-up corresponds to a 7pp decrease in likelihood of take-up by a household (again, suggesting that the substitution effect may dominate). Therefore, the sampled data has lead to severe under-estimation and even sign-switching of

[45]The estimating approach ignores problems raised by a discrete dependent variable, following the approach taken in this literature (e.g., Bramoulle and Kranton, 2007; Gaviria and Raphael, 2001; Sacerdote, 2001). The estimated standard errors handle the heteroskedasticity of the binary response variable.

the endogenous network effect of interest. In particular, partial sampling may cause the researcher to mistake a substituting peer effect for a complementary peer effect. The remainder of the table suggests that, in addition, the exogenous peer effect is also under-estimated and, in the sampled data cases, the effects seem to load on the exogenous own covariate coefficient.

7. Sample Design

We discuss how researchers can adopt our framework to think about data collection. The question we are interested in is: given that a researcher faces a budget constraint and needs to trade off the sampling rate and the number of networks in her sample, is there a method by which she can assess the trade-off?

Suppose that a researcher is interested in estimating a coefficient in a regression of an outcome on a network statistic. Assume that the researcher has a project budget b and a pilot budget p. Each village has a fixed cost f associated with the survey as well as a variable cost c for sampling.[46] We assume that the cost to sample individuals is linear and therefore the cost to sample a ψ-sample of the village is $c\psi n$. Finally, let $\bar{R}$ be the maximum number of villages available to study.

We posit that the researcher is interested in minimizing mean-squared error (MSE) in the estimation of β_0.[47] The relevant program is[48]

$$\min_{\psi\in[0,1],R\leq\bar{R}} \mathrm{MSE}(\psi,R) \text{ s.t. } (c\psi n+f)\,R\leq b. \tag{7.1}$$

At the optimum $\psi=(b/R-f)/(cn)$ and therefore we may consider the concentrated objective function $\mathrm{MSE}(R)=\mathrm{MSE}(\psi(R),R)$. The researcher may estimate the MSE by fully sampling a small number of networks and hypothesizing β_0 and R^2 from the linear regression, in a manner analogous to performing power calculations by positing effect sizes and intra-cluster correlations before conducting a field experiment (e.g., Duflo et al., 2007). A researcher first randomly selects k of the $\bar{R}$ graphs using the pilot budget, where $k=p/(cn+f)$. Then, using these k networks, the researcher conducts a numerical experiment, sampling them at different rates and applying graphical reconstruction to estimate the MSE. By doing this, she can select the optimal ψ and R.

Algorithm (Research Design).

(1) Pick network statistics and network-based hypotheses to test.
(2) Hypothesize β, R^2, and generate outcome variable.
(3) Randomly sample $k=p/(cn+f)$ out of $\bar{R}$ villages and obtain entire networks.
(4) For each $R\in\{\underline{R},...,\bar{R}\}$
    (a) Randomly draw, with replacement, R villages from the collection of k networks.

[46] This method can be applied to richer budgeting frameworks.
[47] Researchers can replace this with an objective function of their choosing.
[48] For formal asymptotics we may have to let $b=b_n$ grow such that $\frac{b_n}{\bar{R}n}\to k$ some positive constant.

(b) Estimate MSE(R) using the sample and hypothesized parameter values.

(i) Sample each of the R village networks at rate $\psi(R) = (b/R - f)/(cn)$.

(ii) Apply graphical reconstruction to estimate $\widehat{\beta}_{\text{ols}}$ using outcome variable from (2).

(iii) Repeat 4(a) and 4(b).i-ii for B simulations.

(5) Pick $R^* \in \text{argmin}_R \widehat{\text{MSE}}(R)$ *and pick* $\psi^* = \psi(R^*)$.

The algorithm enables the researcher to estimate the trade-off she faces, given her interest in specific network effects and the distribution of graphs in her region of study. We conduct a simulation exercise to demonstrate this procedure. We set $b = \$152,400$, $f = \$1200$, $c = \$12$, $n = 200$, $\bar{R} = 150$ and assume that the networks are drawn from the family described in section 5 and the empirical Indian networks.[49] We consider a grid of $R \in \{33, 40, 50, 60, 70\}$ and $\psi \in \{1, 0.7, 0.4, 0.2, 0.1\}$.

Figure 9 displays results for two node-level statistics, eigenvector centrality and clustering, as well as a network-level statistic, the maximal eigenvalue of the adjacency matrix (λ_1). We repeat the exercise for both our simulated network data as well as the Indian networks. The figure shows MSE$(\psi(R), R)$ for sampled networks and graphical reconstruction. It also displays a theoretical lower bound on MSE by plotting the MSE corresponding to using R graphs sampled at 100% instead of at $\psi(R)$. Of course, we find that MSE increases greatly as we move away from 100% sampling and use the raw sampled data. Next we turn to graphical reconstruction and focus on the star subgraph. Looking at the Indian networks, eigenvector centrality has the lowest MSE at 40% sampling while clustering has an optimum at 70%. For these statistics, the simulated networks give 100% as the optimum. Meanwhile, λ_1 has the lowest MSE at 100% sampling with the Indian networks but 20% sampling is the optimum in the simulated networks.

Taken together, the results suggest that, first, performing graphical reconstruction is very important, even with model misspecification as the researcher will not know the true families generating the empirical networks. Second, the MSE-minimizing sampling rate depends greatly on parameters, the network family, and the statistic of interest. It is difficult, if not impossible, to say *ex ante* where the optimum lies: systematic procedures that depend on the setting may be better than rules of thumb. Third, the results push against the prevailing habit of researchers to obtain more cluster-units (e.g., villages) at lower sampling rates when conducting cluster-level analysis. Our results suggest that, at times, just obtaining better data with fewer cluster-units may be worthwhile. Though it is not surprising that network-level statistics exhibit higher levels of MSE, as there are only R as opposed to nR observations, this says nothing about the trade-off between the sampling rate and number of villages.

[49] The numbers are motivated from Banerjee et al. (2013).

8. Conclusion

Applied social network analysis often use graphs constructed from data collected from a partial sample of nodes. Even when nodes are selected randomly, the partial sampling induces non-classical measurement error and consequently biases estimates of regression coefficients and GMM parameters. Moreover, these biases are of unclear sign and magnitude. We analytically examine the biases in the estimation of a number of network-based regression and GMM models with applications to a variety of economic environments. To address the problem in general, we develop a method to construct estimators that are consistent and asymptotically normally distributed using graphical reconstruction, while allowing for substantial heterogeneity across networks. Specifically, the method allows for every network in the sample to be generated by a different model.

We conclude that network-based applied work must proceed cautiously, paying close attention to network data quality. From an applied perspective, researchers should be careful to work either with specifications which provide conservative results when facing sampled data or implement bias correction procedures if possible. Moreover, researchers ought to address the bias problem *ex ante*, either by choosing a unit of study where more complete data is available, using graphical reconstruction to understand how mean-squared error may vary with the sampling rate, or in cases where possible sampling in a way that preserves the properties of the network of economic interest (recognizing that this may not always be possible). Undoubtedly, the performance of graphical reconstruction with empirical network data will only improve as the burgeoning literature on consistently estimable network formation models matures. To that end, from a theoretical perspective the lacuna in the literature is the absence of network formation models that both allow for higher-order dependencies in link formation and are also consistently estimable. This has now become a space of active research.

References

Alatas, V., A. Banerjee, A. G. Chandrasekhar, B. Olken, and R. Hanna (2016): "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia," *American Economic Review*. 1, 2.2, 3.1, 3.1, 4.1, F

Ali, M. M. and D. S. Dwyer (2009): "Estimating peer effects in adolescent smoking behavior: a longitudinal analysis," *Journal of Adolescent Health*, 45, 402–408. F

Ambrus, A., M. Mobius, A. Szeidl, and L. Center (2010): "Consumption risk-sharing in social networks," *NBER Working Paper*. 3.1, 5.2

Andrews, D. (1992): "Generic uniform convergence," *Econometric Theory*, 8, 241–257. C.1

——— (1994): "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica: Journal of the Econometric Society*, 62, 43–72. A.2, C.1

Aral, S. and D. Walker (2011): "Creating social contagion through viral product design: A randomized trial of peer influence in networks," *Management science*, 57, 1623–1639. 2.2

Badev, A. (2016): "Discrete Games in Endogenous Networks: Theory and Policy," *mimeo: University of Pennsylvania*. 1

Bandiera, O. and I. Rasul (2006): "Social Networks and Technology Adoption in Northern Mozambique*," *The Economic Journal*, 116, 869–902. F

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013): "Diffusion of Microfinance," *Science*, 341, DOI: 10.1126/science.1236498, July 26 2013. (document), 1, 1, 2.2, 12, 3.1, 4.1, 5.5, 44, 6, 49, F

Barabasi, A. and R. Albert (1999): "Emergence of scaling in random networks," *Science*, 286, 509. 3.1

Bickel, P. and A. Chen (2009): "A nonparametric view of network models and Newman–Girvan and other modularities," *Proceedings of the National Academy of Sciences*, 106, 21068. 4.4.1

Blume, L., W. Brock, S. Durlauf, and Y. Ioannides (2011): "Identification of social interactions," *Handbook of Social Economics*, 1, 855–966. 3.2

Bollobás, B., C. Borgs, J. Chayes, and O. Riordan (2010): "Percolation on dense graph sequences," *The Annals of Probability*, 38, 150–183. 3.1

Boucher, V. and I. Mourifié (2012): "My friend far far away: Asymptotic properties of pairwise stable networks," *Available at SSRN*, 2170803. 1

Bramoulle, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55. 1, 2.2, 3.2, 24, 25

Bramoulle, Y. and R. Kranton (2007): "Risk-sharing networks," *Journal of Economic Behavior & Organization*, 64, 275–294. 45

Cai, J., A. De Janvry, and E. Sadoulet (2015): "Social networks and the decision to insure," *American Economic Journal: Applied Economics*, 7, 81–108. 1, 12, 3.1, F

Chandrasekhar, A. G. and M. O. Jackson (2016): "A network formation model based on subgraphs," *Available at SSRN*. 1, 4.4.1, 4.4.1, 5.1, 6, A.3.2, C.3

Chatterjee, S., P. Diaconis, and A. Sly (2010): "Random graphs with a given degree sequence," *arXiv Preprint: 1005.1136*. 4.4.1, C.4

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009): "Inference on counterfactual distributions," . E.1

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman (2010): "An Empirical Model for Strategic Network Formation," *NBER Working Paper*. 1, 4.4.1

Chung, F. (1997): *Spectral Graph Theory*, American Mathematical Society. 3.1, 5.2

Conley, T. and C. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, 100, 35–69. 1, F

Currarini, S., M. Jackson, and P. Pin (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77, 1003–1045. 1

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010): "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2, 241–275. 1, 1, 2.2, 3.2, 26

de Paula, A. (2015): "Econometrics of Network Models," *mimeo, to appear in "Advances in Economics and Econometrics, Theory and Applications: Eleventh World Congress of the Econometric Society" Cambridge University Press*. 1

de Paula, A., S. Richards-Shubik, and E. T. Tamer (2014): "Identification of Preferences in Network Formation Games," *Available at SSRN 2577410*. 1

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions," *Quarterly Journal of Economics*, 118, 909–968. 3.1

Diaconis, P. and D. Freedman (1984): "Partial exchangeability and sufficiency," *Statistics: Applications and New Directions*, 205–236. 4.4.1

Dudley, R. (2002): *Real analysis and probability*, vol. 74, Cambridge Univ Pr. E.1

Duflo, E., R. Glennerster, and M. Kremer (2007): "Using randomization in development economics research: A toolkit," *Handbook of Development Economics*, 4, 3895–3962. 7

Fafchamps, M. and S. Lund (2003): "Risk-sharing networks in rural Philippines," *Journal of Development Economics*, 71, 261–287. F

Feigenberg, B., E. Field, and R. Pande (2010): "Building Social Capital Through MicroFinance," *NBER Working Papers*. F

Frank, O. (1980): "Sampling and inference in a population graph," *International Statistical Review/Revue Internationale de Statistique*, 33–41. 1

——— (1981): "A survey of statistical methods for graph analysis," *Sociological methodology*, 12, 110–155. 1

Galeotti, A. and B. W. Rogers (2013): "Strategic immunization and group structure," *American Economic Journal: Microeconomics*, 5, 1–32. 28

Gaviria, A. and S. Raphael (2001): "School-based peer effects and juvenile behavior," *Review of Economics and Statistics*, 83, 257–268. 45

Ghani, A., C. Donnelly, and G. Garnett (1998): "Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases," *Statistics in Medicine*, 17, 2079–2097. 1

Goeree, J., M. McConnell, T. Mitchell, T. Tromp, and L. Yariv (2010): "The 1/d law of giving," *American Economic Journal: Microeconomics*, 2, 183–203. 1, 12, 3.1, F

Goldsmith-Pinkham, P. and G. Imbens (2011): "Determinants of Social Networks: Large Sample Theory," *Harvard Working Paper*. 4.4.1

——— (2013): "Social Networks and the Identification of Peer Effects," *Journal of Business and Economic Statistics*, 31:3, 253–264. 1

Golub, B. and M. Jackson (2009): "How homophily affects learning and diffusion in networks," *Working Paper*. 3.1

——— (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149. 3.1

Graham, B. S. (2014): "An empirical model of network formation: detecting homophily when agents are heterogenous," Tech. rep., National Bureau of Economic Research. 1

Granovetter, M. S. (1973): "The Strength of Weak Ties," *The American Journal of Sociology*, 78, 1360–1380. 1

Hahn, J. and G. Kuersteiner (2004): "Bias reduction for dynamic nonlinear panel models with fixed effects," *Econometric Theory*, 1, 1–40. 6, 1, 58

Hahn, J. and W. Newey (2004): "Jackknife and analytical bias reduction for nonlinear panel models," *Econometrica*, 72, 1295–1319. 6, 1, 4.4, E.1, E.2, E.2

Hall, P. and J. Horowitz (1996): "Bootstrap critical values for tests based on generalized-method-of-moments estimators," *Econometrica: Journal of the Econometric Society*, 891–916. E.2, E.2

Handcock, M. and K. Gile (2010): "Modeling social networks from sampled data," *The Annals of Applied Statistics*, 4, 5–25. 1

Hochberg, Y., A. Ljungqvist, and Y. Lu (2007): "Whom you know matters: Venture capital networks and investment performance," *The Journal of Finance*, 62, 251–301. 1

Iyer, R. and M. Puri (2012): "Understanding bank runs: the importance of depositor-bank relationships and networks," *The American Economic Review*, 102, 1414–1445. 2.2

Jackson, M. (2008a): "Average distance, diameter, and clustering in social networks with homophily," *Internet and Network Economics*, 4–11. 3.1

——— (2009a): "Networks and economic behavior," *Annu. Rev. Econ*, 1, 489–513. 1

——— (2009b): "An Overview of Social Networks and Economic Applications," *The Handbook of Social Economics., ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson. Elsevier Press. forthcoming.* 1

Jackson, M., T. Rodriguez-Barraquer, and X. Tan (2012): "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102, 1857–1897. 3.1

Jackson, M. and B. Rogers (2007a): "Meeting strangers and friends of friends: How random are social networks?" *American Economic Review*, 97, 890–915. 3

——— (2007b): "Relating network structure to diffusion properties through stochastic dominance," *The BE Journal of Theoretical Economics*, 7, 1–13. 1, 3.3, 3.3, 5.4, 5.5, 43, B

Jackson, M. O. (2008b): *Social and Economic Networks*, Princeton University Press. 1, 4.4.1, 5.2, B

Jackson, M. O. and L. Yariv (2007): "Diffusion of behavior and equilibrium properties in network games," *The American economic review*, 97, 92–98. 3.3, 28

Jenish, N. and I. Prucha (2009): "Central limit theorems and uniform laws of large numbers for arrays of random fields," *Journal of Econometrics*, 150, 86–98. 51

Kinnan, C. and R. Townsend (2011): "Kinship and Financial Networks, Formal Financial Access and Risk Reduction," *Northwestern University and MIT Working Paper*. 1, 3.1

——— (2012): "Kinship and financial networks, formal financial access, and risk reduction," *The American Economic Review*, 102, 289–293. F

Kleibergen, F. (2007): "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics," *Journal of Econometrics*, 139, 181–216. 5.3

Kolotilin, A. (2013): "Estimation of a scale-free network formation model," *Available at SSRN 2390446*. 1

Kossinets, G. (2006): "Effects of missing data in social networks," *Social Networks*, 28, 247–268. 1

Kremer, M. and E. Miguel (2007): "The Illusion of Sustainability," *The Quarterly Journal of Economics*, 122, 1007–1065. 1, F

Lahiri, S. N. (1992): "Edgeworth Correction by 'Moving Block' Bootstrap for Stationary and Nonstationary Data," in *Exploring the Limits of Bootstrap*, ed. by R. Lepage and L. Billard, Wiley-Interscience, vol. 270, 183–214. E.2

Lee, L.-f., X. Liu, and X. Lin (2010): "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, 13, 145–176. 24

Leider, S., M. Mobius, T. Rosenblat, and Q. Do (2009a): "Directed Altruism and Enforced Reciprocity in Social Networks*," *Quarterly Journal of Economics*, 124, 1815–1851. F

Leider, S., M. M. Möbius, T. Rosenblat, and Q.-A. Do (2009b): "Directed Altruism and Enforced Reciprocity in Social Networks," *Quarterly Journal of Economics*, 124, 1815–1851. 1, 12, 3.1

Leung, M. (2013): "Two-Step Estimation of Network-Formation Models with Incomplete Information," *mimeo*. 1

——— (2014): "A Random-Field Approach to Inference in Large Models of Network Formation," *Stanford Working Paper*. 1

Ligon, E. and L. Schechter (2012): "Motives for sharing in social networks," *Journal of Development Economics*, 99, 13–26. F

Lin, C., Z. Lin, and C. Lu (1996): *Limit theory for mixing dependent random variables*, vol. 378, Springer. E.2

Macours, K. (2003): "Comparing a Direct with an Indirect Approach to Collecting Household Level data: Who tells the truth about what?" *Washington, DC, United States: Johns Hopkins University. Mimeographed document.* 4.1

Manski, C. (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60, 531–542. 1, 2.2, 3.2

Marmaros, D. and B. Sacerdote (2006): "How do friendships form?" *The Quarterly Journal of Economics*, 79–119. F

Mele, A. (2016): "A Structural Model of Dense Network Formation," *Econometrica, forthcoming*. 1

Menzel, K. (2016): "Strategic network formation with many agents," . 1

Newman, M., S. Strogatz, and D. Watts (2001): "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, 64, 26118. 3.1

Ngatia, M. (2011): "Social interactions and individual reproductive decisions," *Unpublished Working Paper*, 2. 3.2, F

Pastor-Satorras, R. and A. Vespignani (2001): "Epidemic spreading in scale-free networks," *Physical review letters*, 86, 3200. 28

Rothenberg, R. (1995): "Commentary: sampling in social networks," *Connections*, 18, 104–110. 1

Rubin, D. (1976): "Inference and missing data," *Biometrika*, 63, 581. 9

Sacerdote, B. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 681–704. 45

Santos, P. and C. Barrett (2008): "What do we learn about social networks when we only sample individuals? Not much." *Working Paper*. 1, 7, 4.4.1

Shalizi, C. and A. Rinaldo (2012): "Consistency under Sampling of Exponential Random Graph Models," *ArXiv 1111.3054v3*. 4.4

Sheng, S. (2016): "Identification and Estimation of Network Formation Games," *mimeo*. 1

Snijders, T. (1992): "Estimation on the basis of snowball samples: How to weight?" *Bulletin de méthodologie sociologique*, 36, 59. 1

Stephan, F. (1945): "The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate," *The Annals of Mathematical Statistics*, 16, 50–61. 55

Suri, T. (2011): "Selection and comparative advantage in technology adoption," *Econometrica*, 79, 159–209. 4.1

Takasaki, Y., T. Barham Oliver, and L. Bradford (2000): "Rapid rural appraisal in humid tropical forests: an asset possession-based approach and validation methods for wealth assessment among forest peasant households," *World Development*, 28, 1961–1977. 4.1

Thompson, S. (2006): "Adaptive web sampling," *Biometrics*, 62, 1224–1234. 7

Townsend, R. M. (2007): "Townsend Thai Project Household Annual Resurvey, 1998-2006," [Computer file]. 2nd Release. Chicago, IL: Social Sciences Computing Services, the University of Chicago [Producer & distributor]. 4.1

Van der Vaart, A. and J. Wellner (1996): *Weak convergence and empirical processes*, Springer Verlag. C

Watts, D. and S. Strogatz (1998): "Collective dynamics of small-world networks," *Nature*, 393, 440–442. 3.1

Figure 7. Graph-Level Regressions (Simulated Nework Data)

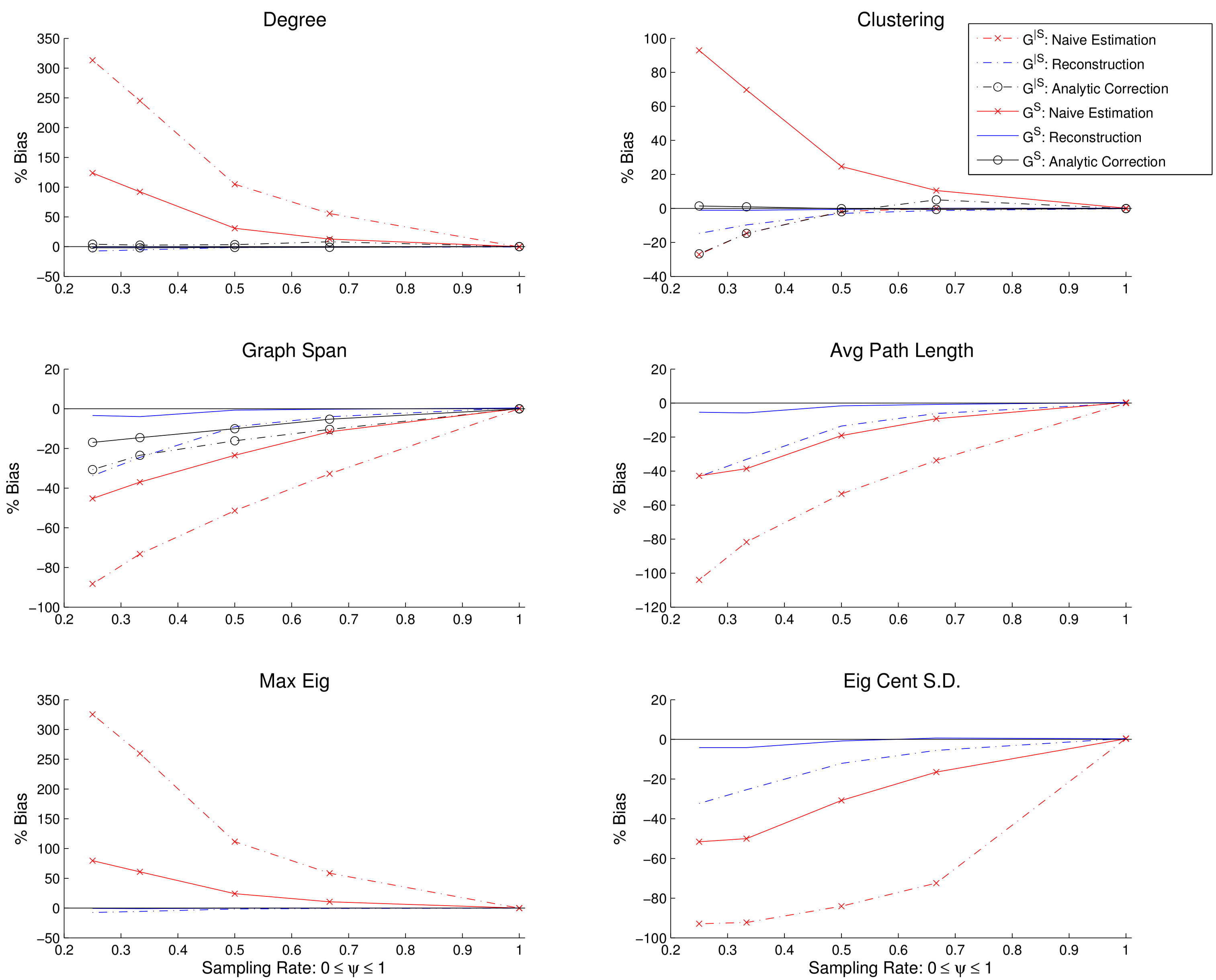


Figure 8. Node-Level Regressions (Simulated Network Data)

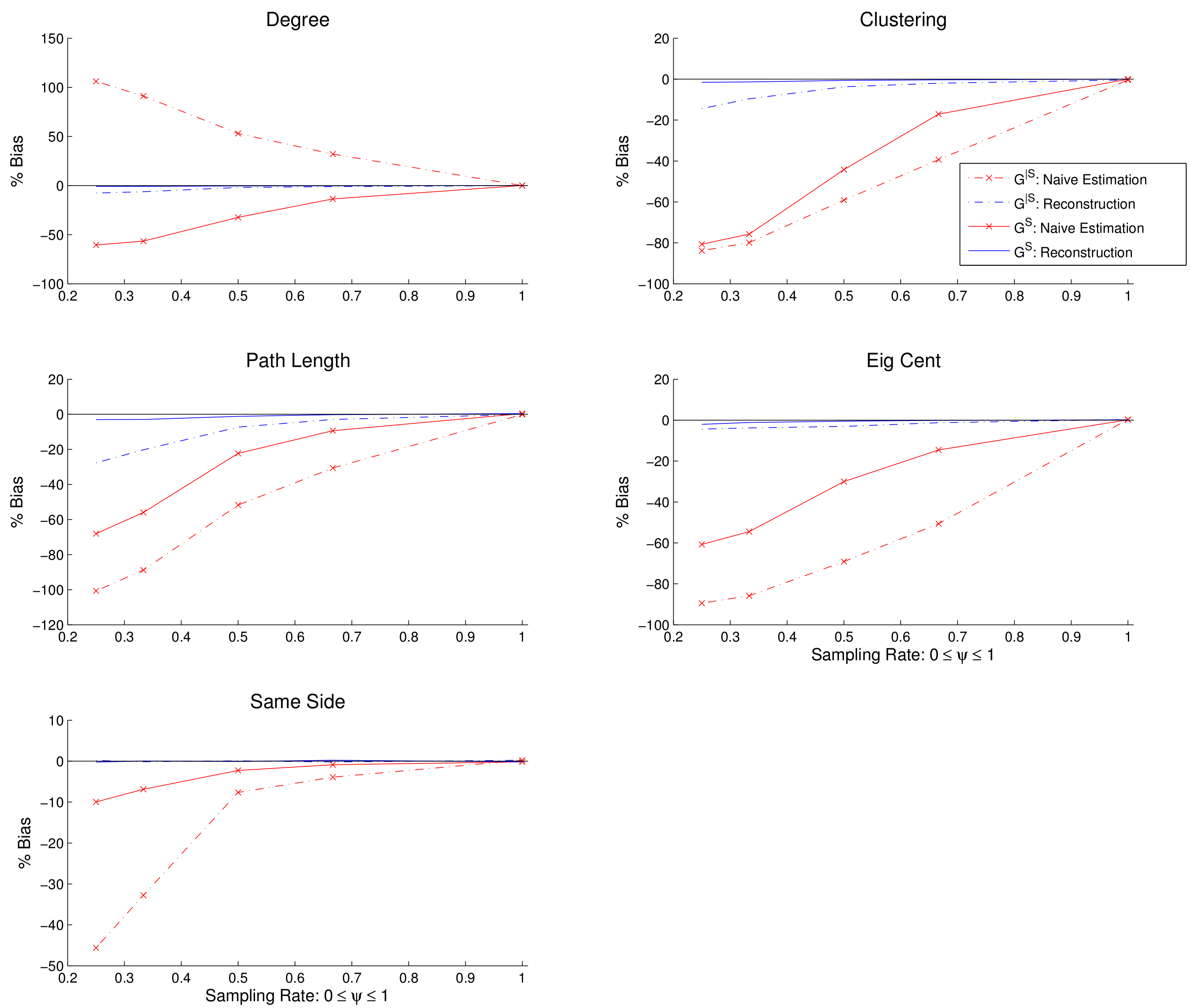


FIGURE 9. Budget Trade-Offs

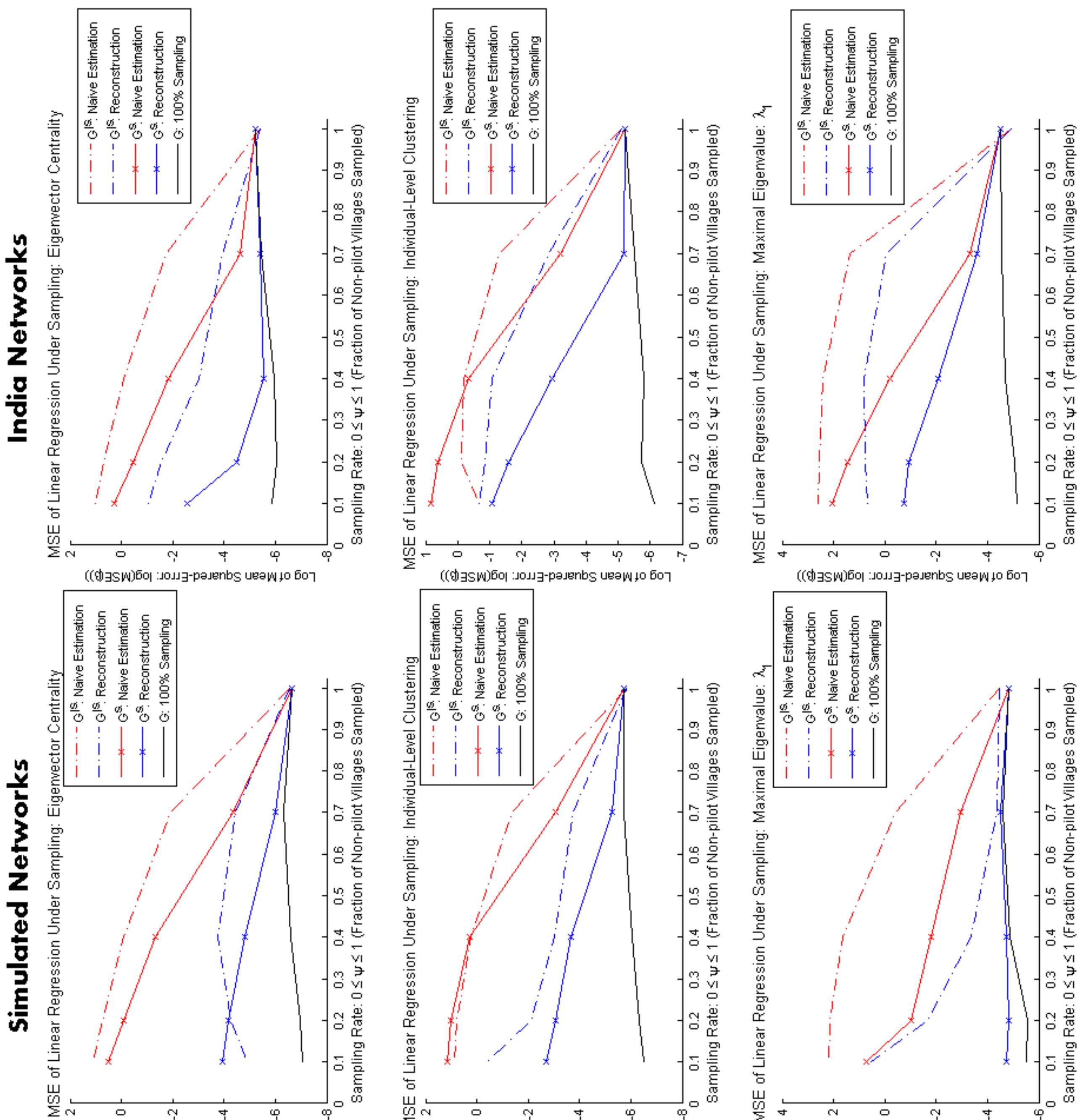


TABLE 1. t-statistic Diagnosis (Simulated Network Data)

	# of Stats where $t_{GR} / t_{Naive} > 1$ out of tota				Mean of $t_{Reconstructed} / t_{Naive}$			
	1/4	1/3	1/2	2/3	1/4	1/3	1/2	2/3
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: Graph Level Regressions								
$G^{\|S}$	6/7	7/7	7/7	7/7	5.32	2.15	1.61	1.29
G^{S}	5/7	7/7	6/7	3/7	1.61	1.41	1.26	1.08
Panel B: Node Level Regressions								
$G^{\|S}$	5/6	5/6	6/6	6/6	4.09	2.74	2.21	1.83
G^{S}	6/6	6/6	6/6	6/6	3.23	2.94	2.46	1.93

Notes: The left panel displays the fraction of times the t-statistic increases when using graphical reconstruction as compared to using the raw sampled network statistic across network statistics. The right panel displays the average ratio of t-statistic under graphical reconstruction to t-statistic under the raw sampled network statistic across network statistics. We use 7 graph-level statistics and 6 node-level network statistics, identical to those used in Tables 1 and 2. The simulation data is the same as that used in Tables 1 and 2.

TABLE 2. Peer Effects Regressions (Simulated Graphs)

	Model 1: 1 Graph; σ_ε=1; 50k Sims.					*Model 2:* 1 Graph; σ_ε=¼; 50k Sims.					*Model 3:* 20 Graphs; σ_ε=1; 2,500 Sims.				
	1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1
Panel A: 2SLS Mean Bias %	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
$G^{\mid S}$ m obs.	1606%	-146%	-56%	-40%	-1%	-107%	-78%	-56%	-37%	0%	-56%	-46%	-32%	-19%	0%
	[1682%]	[54%]	[2%]	[0%]	[0%]	[39%]	[2%]	[0%]	[0%]	[0%]	[1%]	[1%]	[0%]	[0%]	[0%]
G^S (y,x)-census, n obs.	-69%	-55%	-27%	-12%	-1%	-69%	-55%	-27%	-11%	0%	-35%	-22%	-5%	0%	0%
	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]
G^S (y,x)-sample, m obs.	553%	404%	32%	54%	-2%	-81%	900%	288%	-267%	0%	432%	382%	140%	62%	0%
	[776%]	[442%]	[328%]	[51%]	[0%]	[302%]	[1848%]	[145%]	[443%]	[0%]	[147%]	[19%]	[1%]	[1%]	[0%]
G^S Analytic Correction; (y,x)-census, m obs.	18%	-19%	-3%	-2%	-1%	-1%	-1%	0%	0%	0%	-1%	0%	0%	0%	0%
	[178%]	[9%]	[1%]	[0%]	[0%]	[1%]	[0%]	[0%]	[0%]	[0%]	[1%]	[0%]	[0%]	[0%]	[0%]
Panel B: 2SLS Mean Bias %															
$G^{\mid S}$ m obs.	-3692%	70%	-58%	-34%	1%	-30%	-62%	-53%	-36%	0%	-125%	-122%	-102%	-73%	0%
	[3602%]	[111%]	[5%]	[1%]	[1%]	[75%]	[3%]	[1%]	[0%]	[0%]	[1%]	[1%]	[1%]	[1%]	[0%]
G^S (y,x)-census, n obs.	-86%	-80%	-61%	-28%	1%	-83%	-78%	-58%	-28%	0%	-150%	-143%	-104%	-51%	0%
	[1%]	[1%]	[1%]	[1%]	[1%]	[0%]	[0%]	[0%]	[0%]	[0%]	[1%]	[1%]	[0%]	[0%]	[0%]
G^S (y,x)-sample, m obs.	-900%	-905%	-113%	-113%	2%	194%	-1631%	-614%	584%	0%	-913%	-800%	-295%	-129%	0%
	[1486%]	[935%]	[701%]	[108%]	[1%]	[606%]	[3642%]	[321%]	[961%]	[0%]	[304%]	[39%]	[3%]	[1%]	[0%]
G^S Analytic Correction; (y,x)-census, m obs.	-41%	38%	5%	2%	1%	2%	2%	0%	0%	0%	2%	0%	0%	0%	0%
	[361%]	[19%]	[1%]	[1%]	[1%]	[2%]	[1%]	[0%]	[0%]	[0%]	[1%]	[1%]	[1%]	[0%]	[0%]
Panel C: Concentration Parameter															
$G^{\mid S}$ m obs.	1.2	1.8	3.7	6.6	15.6	15.7	26.6	60.9	111.6	272.0	17	28	51	74	111
G^S (y,x)-census, n obs.	7.0	8.2	10.6	13.1	15.6	118.2	139.6	182.1	227.3	271.2	131	136	128	119	112
G^S (y,x)-sample, m obs.	1.1	1.2	1.9	4.2	15.5	15.2	16.5	28.7	69.6	271.0	1	3	14	36	112
G^S Analytic Correction; (y,x)-census, m obs.	1.7	2.9	6.3	10.0	15.5	26.1	47.2	106.4	172.2	271.6	25	39	70	94	112

Estimators: The model is $y = \alpha + \rho_0 w(G)y + \gamma_0 x + \delta_0 w(G)x + \varepsilon$. We consider four estimators. Each estimator uses a sampled graph, a sample of (y,x), and a data sample for the 2-stage least squares (2SLS) estimation. The sampled graph is either $G^{|S}$ or G^S, the sample of (y,x) is either for the sampled nodes or for an entire census, and the data sample for 2SLS is either data for the m sampled nodes or all available data (any unknown or missing data is replaced with zeros) for the census of n nodes. The "Analytic Correction" uses G^S, the census of (y,x), and 2SLS estimation on only the data for the m sampled nodes. The concentration parameter c is given as follows following Kleibergen (2007). For a first stage $X = Z\pi + v$, we use $c = \pi' S^{-1} \pi$ where $S := var(p)$ and $p = (Z'Z)^{-1}Z'X$.

Notes: Estimation is performed using homophily graphs with 250 nodes, 6 cliques of equal size, and expected degree of ~15. The simulation model parameters are ρ0=0.5 and δ0=0.5. All bias percentages are computed relative to these parameter values. Standard errors of the simulation means are shown in brackets. For comparison, OLS biases at 100% sampling are [ρ0; δ0] = [46%; -98%], [8%; -16%], and [28%; -63%], reading across the three columns.

Table 3. Bias in Estimation of β_0 in Jackson and Rogers (2007a) Model

	Raw Network Data				*Graphical Reconstruction*			
	1/4	1/3	1/2	2/3	1/4	1/3	1/2	2/3
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: Simulated Networks, Bias % in Estimation of β_0								
$G^{\mid S}$	329.0%	250.0%	104.0%	55.0%	1.0%	0.0%	0.0%	0.0%
G^S	117.0%	85.0%	28.0%	12.0%	0.0%	0.0%	0.0%	0.0%
Panel B: Indian Networks, Bias % in Estimation of β_0								
$G^{\mid S}$	263.0%	230.0%	103.0%	53.0%	8.0%	8.0%	7.0%	6.0%
G^S	129.0%	92.0%	31.0%	14.0%	6.0%	5.0%	3.0%	1.0%

Notes: Table presents bias in estimation of β_0 the transmission parameter in the Jackson-Rogers diffusion model described in section 3.3 of the text. Data generating process for the simulated networks is the same as in Table 2. Networks in panel B are same as those described in Table 6. We set β_0 to exp(-2) and perform each simulation 100 times.

Table 4. Summary of Results from Numerical Simulations with Indian Village Networks

	Network Level Regressions				*Node Level Regressions*			
	1/4	1/3	1/2	2/3	1/4	1/3	1/2	2/3
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: G^S								
Raw Network Statistic Median Bias	67.52%	48.62%	23.48%	9.88%	88.02%	70.10%	48.68%	25.33%
Reconstruction Median Bias	0.90%	0.56%	0.35%	0.10%	6.97%	4.62%	1.84%	0.73%
Reconstruction Reduces Bias	8/8	8/8	8/8	8/8	5/5	5/5	5/5	5/5
Raw Network Statistic Bias	67.52%	47.76%	23.48%	11.56%				
Analytic Correction Median Bias	14.58%	6.94%	2.33%	0.65%				
Analytic Corrections Reduce Bias	3/3	3/3	3/3	3/3				
Panel B: $G^{\mid S}$								
Raw Network Statistic Median Bias	88.27%	73.20%	51.36%	32.78%	90.94%	81.50%	60.25%	38.46%
Reconstruction Median Bias	13.50%	5.79%	2.28%	0.66%	28.43%	16.91%	7.98%	3.97%
Reconstruction Reduces Bias	7/8	8/8	8/8	8/8	5/5	5/5	5/5	5/5
Raw Network Statistic Bias	190.26%	132.95%	75.19%	41.41%				
Analytic Correction Median Bias	18.57%	13.62%	8.69%	5.41%				
Analytic Corrections Reduce Bias	2/2	2/2	2/2	2/2				

Notes: We ran all the regressions from Tables 1 and 2. We present the median bias across several network statistics at each sampling rate, where the network statistics are those analyzed in Tables 1 and 2. Networks are drawn from the data-set of Banerjee et al. (2011) where the graphs are treated as full networks and then sampled at rates 1/4, 1/3, 1/2, and 2/3. The network formation model used is a conditional edge-independence logistic regression model where covariates are distance in GPS coordinates, number of beds, number of rooms, number of latrines, and electricity provision. The full table is available upon request.

TABLE 5. Empirical Application: Diffusion of Microfinance

Panel A: Graph Level Regressions of Microfinance Take-up Rate of Village on Network Characteristics

	Eigenvector Centrality of Injection Point	Average Path Length	Conductance	Var. of Eig. Centrality Distribution
	[1]	[2]	[3]	[4]
G^S	1.6335	-0.0545	0.9659	58.6554
	[0.8715]	[0.0488]	[0.2671]	[40.5225]
Graphical Reconstruction	2.4244	-0.0928	1.4443	87.7406
	[1.0680]	[0.0597]	[0.3644]	[49.4477]

Panel B: Vertex Level Regressions of Microfinance Take-up Decision of a Household on Network Characteristics

	Eigenvector Centrality of Household	Inverse Distance to Injection Point
	[1]	[2]
G^S	0.5528	0.0221
	[0.1887]	[0.0277]
Graphical Reconstruction	0.662	0.0453
	[0.2091]	[0.0286]

Panel C: Regression of Household's Take-up Decision on Neighbors' Take-up Decisions

	ρ	γ	δ
	[1]	[2]	[3]
G^S, (y,x)-Census; n obs.	-0.027274	0.058066	0.016371
	[0.0221]	[0.0138]	[0.0115]
G^S, (y,x)-Sample; m obs.	0.047085	0.079207	-0.00090318
	[0.0098]	[0.0175]	[0.0047]
Analytic Correction, (y,x)-Census; m obs.	-0.071297	0.025769	0.025271
	[0.049]	[0.0206]	[0.0163]

Notes for Panel A: This panel provides estimates of regressions of the microfinance take-up rate of a village on its network characteristics. Each column presents a separate regression and all regressions control for number of households, fraction of households with savings access, and fraction of households that are SC/ST (scheduled caste/ scheduled tribe). The sample consists of 43 villages. The mean of the dependent variable is 0.1849. All regressions control for the number of nodes in the graph. Standard errors are heteroskedasticity robust.

Notes for Panel B: This panel provides estimates of regressions of whether a household decides to join microfinance on its network characteristics. The sample consists of 8441 households. The mean of the dependent variable is 0.1750. All regressions include village fixed effects. Standard errors are clustered by village. Inverse distance to injection point is the inverse of the minimal path length from the household in question to the set of initially informed households.

Notes for Panel C: This panel provides 2SLS estimates of regressions of whether a household joins microfinance on whether its neighbors join microfinance. The sample of (y,x) is either for just the sampled nodes or the entire census and the data sample for 2SLS is either data for the m sampled nodes or all available data for the census of n nodes. The analytic correction uses the induced subgraph, the census of (y,x), and 2SLS estimation on only the data for the m sampled nodes. Standard errors are clustered by village.

Appendix A. Regularity Conditions

A.1. Regularity Conditions for Section 3.

Assumption A.1. *Let (a_R) be a sequence of normalizing constants.*

(1) *The data consists of $(y_{rR}, X_{rR})_{r\leq R}$ with $\mathrm{E}\left[\epsilon_R | X_R\right] = 0$ and the model is $y_{rR} = \alpha + X_{rR}\beta_0 + \epsilon_{rR}$.*
(2) $0 < \sigma_x^2 < \infty$ *where* $\sigma_x^2 := \mathrm{plim}_{R\to\infty} a_R^{-2}\mathbb{E}_R\left(X_r - \bar{X}_{Rr}\right)^2$.
(3) $\sigma_v^2 := \mathrm{plim}_{R\to\infty} a_R^{-2}\mathbb{E}_R v_r^2 < \infty$.
(4) *The mismeasured regressor is $X^*_{rR} = \pi_R X_{rR} + v_{rR}$, with $\lim_{R\to\infty}\pi_R = \pi$, and $\mathrm{plim}_{R\to\infty} a_R^{-2}\mathbb{E}_R\left[v_r X_r\right] = 0$.*

Assumption A.2. *The sequence of networks $\{(G_{rR}) : \ r = 1, .., R; \ R \geq 1\}$ is such that $d\left(G_{rR}\right)/a_{1n} \xrightarrow{\mathrm{P}} c_1$ and $d_2\left(G_{rR}\right)/a_{2n} \xrightarrow{\mathrm{P}} c_2$ for constants $a_{1n}, a_{2n} \in o\left(n_R\right)$.*

A.2. Regularity Conditions for Section 4.

Let $\mathrm{P}(G_r|z_r;\theta_r)$ be the distribution of the graph G_r given covariates z_r.

Assumption A.3 (Random Graph Model and First Stage Estimation).

(1) $\forall r$, Θ_r *is a compact subset of* $\mathbb{R}^{d_\theta}$; G_r *is a* $\mathcal{G}_n$*-valued random graph with* $\mathrm{P}(G_r|z_r;\theta_r) \in C^2(\Theta_r)$ *at every* $(G, z) \in \mathcal{G}_n\times\mathcal{Z}$; $\bar{H}_{r,R} := \sup_z \max_{G,\theta_r}\left|\frac{\partial}{\partial\theta}\mathrm{P}(G|z;\theta)\right|$, $\sup_R \sup_r \bar{H}_{R,r} < \infty$.
(2) *The first stage estimation satisfies for some sequence of normalizing constants (a_R), $b > 1$, and $r \leq R$, $a_R\cdot(\widehat{\theta}_r - \theta_{0r}) = O_{\mathrm{P}}(1)$ and $\sup_{r\leq R}\|\widehat{\theta}_r - \theta_{0r}\| = O_{\mathrm{P}}(a_R^{-1}\cdot R^{1/b})$.*
(3) *For node-level analysis $a_R^{-1}\cdot\sqrt{nR^{1+2/b}} \to 0$ and graph-level analysis $a_R^{-1}\cdot\sqrt{R^{1+2/b}} \to 0$.*
(4) *β_0 is an interior point of $\mathcal{B}$, a compact subset of $\mathbb{R}^{d_\beta}$.*

Condition 1 ensures that the random graph family is smooth enough in the parameter, so small deviations from the true parameter do not result in very different probability distributions. Condition 2 is a high-level condition on the first stage estimation which we microfound in section 4.4. It guarantees that we can uniformly replace the estimated network formation parameter for every graph in the sequence with its true value. Condition 3 is a rate requirement which relates the rate of estimation of the network formation process to the rate of estimating the economic model of interest. Condition 4 is a standard interiority condition.

Let h denote a random variable.

Definition A.1. *A sequence of measurable (potentially matrix-valued) functions $\{\phi_{i,r}(h_{ir};\alpha) : i = 1, .., n_R, \ r = 1, ..., R\}$ satisfies an* envelope condition *over $\alpha \in \mathcal{A}$ if there exist measurable functions $L_{i,r}(h_{ir})$, with $\|\phi_{i,r}(h_{ir};\alpha)\| \leq L_{i,r}(h_{ir})$ for every h_{ir} and α, and $\sqrt{n}\mathbb{E}_n L_{i,r,R}$ has uniformly integrable vth moment for $v \geq 2$.*

Definition A.2. *A sequence of measurable (potentially matrix-valued) functions* $\{\phi_{i,r}(h_{ir};\alpha) : i = 1,..,n_R,\ r = 1,...,R\}$ *is* Lipschitz continuous *in* $\alpha \in \mathcal{A}$ *if there exist measurable functions* $M_{i,r}(h_{ir})$ *with* $\|\phi_{i,r}(h_{ir};\alpha) - \phi_{i,r}(h_{ir};\bar{\alpha})\| \le M_{i,r}(h_{ir})\,\|\alpha - \bar{\alpha}\|$ *for every* h_{ir} *and* $\alpha, \bar{\alpha} \in \mathcal{A}$*, and* $\sqrt{n}\mathbb{E}_n M_{i,r,R}$ *has uniformly integrable* v*th moment for* $v \ge 2$.

We also use $\mathcal{I}_{h|x}(\alpha) := \mathrm{E}\left[\frac{\partial}{\partial\alpha'}\log f(h|x;\alpha)\frac{\partial}{\partial\alpha}\log f(h|x;\alpha)|x;\alpha\right]$ to denote the conditional information matrix with random variable $h|x$, density or pmf $f(h|x)$, and parameter α.

Turning to the economic model, observe that the network statistic $w(G)$ may be growing or shrinking in n. For instance, the eigenvector centrality declines as it is a unit norm object. The degree of a node may be $\Theta(1)$, $\Theta(\log n)$, or $\Theta(n)$ depending on the graph family. In what follows, in regression we assume that the model is such that all regressors are rescaled at the appropriate rate: if they exhibit growth or shrinkage at b_n, we assume that the models are specified using $\tilde{w} := b_n^{-1}w$ as regressors. Similarly, in GMM we assume that the moments and network statistics, both of which may depend on R, are appropriately rescaled. We present regularity conditions for least squares and GMM separately, though they essentially can be nested since least squares does not require assuming the joint distribution of y and G.

Assumption A.4 (Linear Regression).

(1) $\mathrm{E}[\epsilon|w] = 0$, $\mathrm{E}[\epsilon\epsilon'|w] = \Omega$, *p.d. with* $\sup_R \lambda_{\max}(\Omega) < \infty$
(2) $\mathrm{E}\left[\|w_{ir}\|^k|x_r;\theta_r\right]$ *and* $\left\|\mathcal{I}_{w_{ir}|x_r}(\theta_r)\right\|$ *for* $k = 1,2$ *satisfy the envelope condition with* $L_{i,r}(x_r)$.
(3) $\sup_{R\ge1}\sup_{r\le R}\mathrm{var}\left(\sqrt{n}\mathbb{E}_n w_{ir}(G_r)\right) < C_1 < \infty$ *and* $\inf_{R\ge1}\inf_{r\le R}\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\theta_{0r})\right) > C_0 > 0$*, uniformly over the array.*

Define $g_R(\beta) := \mathbb{E}_{n,R}\mathrm{E}m(y_{ir}, w_{ir}(G_r);\beta)$ and $f(m|x;\beta,\theta)$ be known up to parameters.

Assumption A.5 (GMM).

(1) $\widehat{W} = W + o_{\mathrm{P}}(1)$*,* W *is p.s.d. and the model satisfies* $\lim_{R\to\infty} W g_R(\beta) = 0$ *only if* $\beta = \beta_0$.
(2) *The limits* $\lim_{R\to\infty}\mathbb{E}_{n,R}[\mathrm{E}\mathcal{E}_{ir}(x_r;\theta_r,\beta)]$ *and* $\lim_{R\to\infty}\mathbb{E}_{n,R}\left[\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\theta_r,\beta)\right]$ *exist uniformly over* $\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r$.
(3) $\left\|\mathcal{I}_{m_{ir}|x_r}(\beta,\theta_r)\right\|$, $\mathrm{E}\left[\left\|\frac{\partial}{\partial\beta'}m(X_{ir};\beta)\right\||x_r;\theta_r,\beta'\right]$, *and* $\mathrm{E}\left[\|m(X_{ir};\beta)\|^k|x_r;\theta_r,\beta'\right]$ *for* $k = 1,2$ *satisfy the envelope condition with envelope* $L_{i,r}(x_r)$.
(4) $m(X;\beta)$ *is continuously differentiable on the interior of* $\mathcal{B}$ *for every* $X \in \mathcal{X}$ *and both* $m(X;\beta)$ *and* $\frac{\partial}{\partial\beta'}m(X;\beta)$ *satisfy the Lipschitz condition with constant* $M_{i,r}(X_{ir})$*, where* $\mathrm{E}[M_{i,r}(y_{ir},w_{ir})|x_r] \le L_{i,r}(x_r)$.
(5) $\sup_{R\ge1}\sup_{r\le R}\mathrm{var}\left(\sqrt{n}\mathbb{E}_n[m(X_{ir};\beta_0)]\right) < C_1 < \infty$ *and* $\inf_{R\ge1}\inf_{r\le R}\mathrm{var}\left(\sqrt{n}\mathbb{E}_n[\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})]\right) > C_0 > 0$*, uniformly over the array.*

Assumptions A.4 and A.5 are similar, and not particularly restrictive, so we discuss the GMM case. Assumption A.5.1 is a standard identification condition. Assumption A.5.2 is standard (e.g., Andrews, 1994) and Assumption A.5.3 places uniform restrictions on higher moments of the conditional moment, slope of the moment, and information matrix allowing weak laws of large numbers to be applied. Assumption A.5.4 allows these convergences to be uniform over the parameter space.

Assumption A.5.5 is what allows us to pass a central limit theorem to the conditional random variable if the unconditional satisfies one.[50] It is reasonable in practice because we use independence across graphs and simply a uniform boundedness condition within graph. This is substantially weaker than having to assume a within-graph central limit theorem for m_{ir}, which would depend on the idiosyncrasies of the network formation model and network statistics. However, it comes at the cost of requiring data from multiple networks. We make this assumption because currently the statistics of networks literature has not characterized within-graph node characteristic interdependencies (e.g., the correlation of eigenvector centrality between nodes for various random graph families).

A.3. **Regularity Conditions for Section 4.4 (Examples).**

A.3.1. *Example 1: Conditional Edge Independence Model.* First define a distance (pseudo-metric) $d_\Xi(\cdot,\cdot)$ over the set of pairs, given by Ξ, where two pairs ij and kl's distance is said to be the minimum coordinate-wise distance between an element of the first pair and an element of the second pair. Specifically, for $s_{ij}, s_{kl} \in \Xi$ and $d_\Lambda(t_i,t_j) := \|t_i - t_j\|_1$,

$$d_\Xi(s_{ij}, s_{kl}) := d_\Lambda(t_i,t_j) \wedge d_\Lambda(t_i,t_l) \wedge d_\Lambda(t_j,t_k) \wedge d_\Lambda(t_k,t_l).$$

To describe interdependence in the data, we define a mixing coefficient. Let $D \subset \mathbb{Z}^d$ be an integer lattice and to each $s \in D$ We associate a random variable z_s. Then $\{z_s : s \in D\}$ is a random field and we are interested in controlling the dependence of z_s and $z_{s'}$. Let $\mathcal{A}_\Omega$ be the σ-algebra generated by a random field $\{z_s : s \in \Omega\}$. We define the mixing coefficient

$$\alpha_{k,l}(m) := \sup\left\{\left|\mathrm{P}\left(A_1 \cap A_2\right) - \mathrm{P}\left(A_1\right)\mathrm{P}\left(A_2\right)\right| : A_i \in \mathcal{A}_{\Omega_i}, |\Omega_1| \le k, |\Omega_2| \le l, d\left(\Omega_1,\Omega_2\right) \ge m\right\}$$

where $d(\Omega_1,\Omega_2) = \min_{x,y\in\Omega_1\times\Omega_2} \|x-y\|_1$. We will need to assume that the level of interdependence goes to zero as the distance between the two subsets goes to infinity.[51]

[50] Since this paper focuses on the effect of sampling on network analysis and not on regression or GMM models on graphs, we make the assumption that the underlying model satisfies reasonable regularity conditions if the full networks were observed and focus on the effect of sampling and graphical reconstruction.

[51] The triangular array notation is cumbersome, see Jenish and Prucha (2009), but formally is $\{z_{s,R} : s \in D_R,\ R \in \mathbb{R}\}$ a triangular array defined on a sequence of probability spaces where D_R is a finite subset of D and

$$\alpha_{k,l}(m) := \sup_{R\ge 1}\sup\left\{\left|\mathrm{P}^R\left(A_1 \cap A_2\right) - \mathrm{P}^R\left(A_1\right)\mathrm{P}^R\left(A_2\right)\right| : A_i \in \mathcal{A}^R_{\Omega_i},\ \left|\Omega_1^R\right| \le k,\ |\Omega_2| \le l,\ \Omega_i \subset D_R,\ d\left(\Omega_1,\Omega_2\right) \ge m\right\}.$$

Assumption A.6 (Mixing Conditions). *$\forall r$, $z_r := \left\{z_{ir} : t_i \in \Lambda \subset \mathbb{Z}^d\right\}$ is a stationary mixing random field, $z_{ijr} := f(z_{ir}, z_{jr})$ satisfies $\sup_r \mathrm{E}\|z_{ijr}\|^{p+\delta} \lesssim \sup_r \mathrm{E}\|z_{1r}\|^{p+\delta}$, and $\sup_r \mathrm{E}\left\|z_{1r}\right\|^{p+\delta} < \infty$ for $p > 2$ with (i) $\sup_r \alpha^r_{2,\infty}(m) \leq Ca^m$ for $a \in (0,1)$ or (ii) $\sup_r \alpha^r_{2,\infty}(m) = o(m^{-d})$.*

The assumption on f is not very restrictive. The most natural example is a covariate based on the difference in characteristics of nodes i and j: $z_{ij} = \|z_i - z_j\|$. It is easy to see that $\mathrm{E}\|z_{ij}\|^k \leq 2^k \mathrm{E}\|z_1\|^k$ by the binomial theorem and stationarity. Because A_{sr} is a measurable function of z_{sr}, it will inherit stationarity and mixing properties and therefore so will X_{sr}.

Assumption A.7 (Joint Convergence). Let $Q_{(r)}(\theta_r) := \mathrm{plim}_{n\to\infty} |\Xi|^{-1} \sum_{s\in\Xi} \mathrm{E}q(X_s;\theta_r)$.

1. $\forall \eta > 0$, $\inf_{r\leq R}\left[Q_{(r)}(\theta_{0r}) - \sup_{\theta:\|\theta-\theta_{0r}\|>\eta} Q_{(r)}(\theta)\right] > 0$.
2. *$D^{|v|}q(X_{sr};\theta)$ satisfies a Lipschitz condition with $B(X_{sr})$, for multi-index $|v| \geq 2$.*
3. *2^{b-1} moments exist for envelope $B(z_{sr}) \geq \|\partial Q_{(r)}(\theta_r)/\partial\theta_r\|$.*
4. *$R = O(|\Xi|^h)$ with $h < p/2 - \gamma dp - 1$ for some $\gamma \in (0,1)$*

Condition 1 is standard for identification, 2 requires sufficient smoothness, and 3 requires that the score functions have well-behaved envelopes. If the link function is logistic, this is satisfied. Condition 4 relates the number of networks to the number of nodes per network: it needs to grow slow enough to be able to uniformly control the estimation error.

A.3.2. *Example 2: Links and Triangles SUGM.*

Assumption A.8 (Links and Triangles SUGM). *Let the probability of links and triangles be $(p_{0L,r}, p_{0T,r}) = \left(\frac{\theta_{0L,r}}{n^{h_L}}, \frac{\theta_{0LTr}}{n^{h_T}}\right)$, where*

1. $(\theta_{0L,r}, \theta_{0T,r}) \in \left[\underline{D}, \overline{D}\right]^2 \subset \mathbb{R}^2_+$,
2. *$h_L \in (1/2, 1]$ and $h_T \in (h_L + 1, \min\{3, 3h_L\})$ or $h_L \in (1,2)$ and $h_T \in [h_L + 1, \min\{3, 3h_L\})$,*
3. *and $R = n^k$ for $k < \max\{2 - h_L, 3 - h_T\}$.*

This first two are just the assumptions required to get consistent and asymptotically normally distributed parameter estimates in Chandrasekhar and Jackson (2016), Proposition 4. The third will allow us to maintain a slow enough rate to get a uniform convergence of the objective functions across the R networks.

Appendix B. Proofs for Section 3

Proof of Lemma 1. This follows from $\sigma_{\widetilde{w}} \cdot \mathrm{plim}\,\widehat{\beta} = \sigma_{\widetilde{w}} \cdot \beta_0 \frac{\mathrm{cov}(\widetilde{w},w)}{\mathrm{var}(\widetilde{w})} \leq \beta_0\sigma_w$, using Cauchy-Schwarz since $|\mathrm{cov}(\widetilde{w}, w)| \leq \sigma_w\sigma_{\widetilde{w}}$ and $\mathrm{cov}(\widetilde{w}, w) > 0$. ■

Lemma B.1. *Under Assumption A.1, $\widehat{\beta} \xrightarrow{\mathrm{P}} \pi^{-1}\beta_0 \frac{\sigma_x^2}{\sigma_x^2+\pi^{-2}\sigma_v^2}$.*

Proof. The proof is standard and follows from $\mathrm{plim}\,\widehat{\beta} = \mathrm{plim}\left(a_R^{-2} X_R^{*\prime} X_R^*\right)^{-1} a_R^{-2}\left(\pi_R X_R + v_R\right)' X_R\beta_0 = \frac{\beta_0}{\pi}\frac{\sigma_x^2}{\sigma_x^2+\pi^{-2}\sigma_v^2}$. ■

Proof of Proposition 3.1. For degree, from Lemma D.1 and D.2,

$$\mathrm{E}[d(G^{|S}_{rR})|G_{rR}] = \underbrace{(\psi + \Theta(n^{-1}))}_{\pi_R}\underbrace{d(G_r)}_{X_{rR}} \text{ and } \mathrm{E}[d(G^{S}_{r})|G_r] = \underbrace{(1-(1-\psi)^2 + \Theta(n^{-1}))}_{\pi_R}\underbrace{d(G_r)}_{X_{rR}}.$$

Similarly, for graph clustering, from Lemma D.3 and D.4,

$$\mathrm{E}\left[\frac{\rho(G^S)}{\tau(G^S)}|G_{rR}\right] = 1\cdot\frac{\rho(G)}{\tau(G)} + o(1). \text{ and } \mathrm{E}\left[\frac{\rho(G^S)}{\tau(G^S)}|G_{rR}\right] = \frac{\psi(3-2\psi)}{1+\psi(1-\psi)}\cdot\frac{\rho(G)}{\tau(G)} + o(1).$$

Finally, for support, from Lemma D.5,

$$\mathrm{E}\left[s\left(G^S\right)|G\right] = \frac{\psi^2 + 2\psi\left(1-\psi\right)\left\{1-\sum_x\left(1-\psi\right)^x \mathrm{P}\left(x|G\right)\right\}}{\psi^2 + 2\left(1-\psi\right)\psi} s\left(G\right) + o\left(1\right),$$

and

$$\mathrm{E}\left[s\left(G^{|S}\right)|G\right] = \left\{1 - \sum_x \left(1-\psi\right)^x \mathrm{P}\left(x|G\right)\right\} s\left(G\right) + o\left(1\right).$$

So Assumption A.1 holds so the result follows from Lemma B.1. ■

Proof of Proposition 3.2. Recall $\zeta_r = d_2(G_r)/d(G_r)$ and let $\log\widetilde{\zeta}_r := \log\zeta_r + \log\gamma$ which we can write by Lemma D.6. To sign the bias we are interested in

$$\lim_{R\to\infty}\frac{R^{-1}\sum\mathrm{cov}\left(\log^{-1}(d_2(\widetilde{G}_r)d(\widetilde{G}_r)^{-1}),\log^{-1}\zeta_r\right)}{R^{-1}\sum\mathrm{var}\left(\log^{-1}(d_2(\bar{G}_r)d(\widetilde{G}_r)^{-1})\right)}.$$

First observe that

$$R^{-1}\sum\mathrm{cov}\left(\log^{-1}(d_2(\widetilde{G}_r)d(\widetilde{G}_r)^{-1}),\log\zeta_r\right) = R^{-1}\sum\mathrm{cov}\left(\log^{-1}\widetilde{\zeta}_r,\log^{-1}\log\zeta_r\right) + o_{\mathrm{P}}(1).$$

This follows from $\left|\log^{-1}(d_2(\widetilde{G}_r)d(\widetilde{G}_r)^{-1}) - \log^{-1}\bar{\zeta}_r\right| = o_{\mathrm{P}}(1)$ which we can see by considering the numerator of the fraction and noting

$$\left|\log\widetilde{\zeta}_r - \log(d_2(\widetilde{G}_r)d(\widetilde{G}_r)^{-1})\right| \leq \left|\log d_2(G_r) - \log d(G_r) + \log\gamma - \log d_2(\widetilde{G}_r) + \log d(\widetilde{G}_r)\right| = o_{\mathrm{P}}(1),$$

by Lemma D.6, where $\gamma = k(\psi)/\psi$ or ψ depending on G^S or $G^{|S}$, using the fact that $\log(\cdot)$ is Lipschitz on $\mathbb{R}_{\geq 1}$. Therefore we are interested in $\lim_{R\to\infty}\frac{R^{-1}\sum\mathrm{cov}\left(\log^{-1}\widetilde{\zeta}_r,\log^{-1}\zeta_r\right)}{R^{-1}\sum\mathrm{var}\left(\log^{-1}\widetilde{\zeta}_r\right)}$.

If $\gamma > \zeta_r^{-1}$ for every r, then $\mathrm{cov}\left(\log^{-1}\widetilde{\zeta}_r,\log^{-1}\zeta_r\right)$ is positive for every r by definition. In addition, for every r, since $\log\gamma < 0$,

$$\mathrm{cov}\left((\log\zeta_r + \log\gamma)^{-1},\log^{-1}\zeta_r\right) < \mathrm{var}\left(\log^{-1}\zeta_r\right).$$

This shows $\lim_{R\to\infty}\frac{R^{-1}\sum\mathrm{cov}\left(\log^{-1}\widetilde{\zeta}_r,\log^{-1}\zeta_r\right)}{R^{-1}\sum\mathrm{var}\left(\log^{-1}\widetilde{\zeta}_r\right)} < 1$.

Next, assume that $\gamma < \zeta_r^{-1}$. Then $sign\left\{cov\left(\log^{-1}\widetilde{\zeta}_r, \log^{-1}\zeta_r\right)\right\}$ depends on the distribution of ζ_r; it cannot be signed. Therefore $\operatorname{plim} R^{-1}\sum \operatorname{cov}\left(\log^{-1}\bar{\zeta}_r, \log^{-1}\zeta_r\right)$ can take either sign. This is easily seen geometrically.

Finally, that the analytical corrections are consistent follows from the above argument in the first step, noting that we use $\log(\gamma d_2(\bar{G}_r)d(\bar{G}_r)^{-1})$ and therefore

$$\left|\log\zeta_r - \log(\gamma d_2(\widetilde{G}_r)d(\widetilde{G}_r)^{-1})\right| \leq \left|\log\widetilde{\zeta}_r - \log(d_2(\bar{G}_r)d(\widetilde{G}_r)^{-1})\right| + |\log\gamma - \log\gamma| = o_{\mathrm{P}}(1)$$

which completes the result. ■

Proof of Proposition 3.3.

Step 1: $G^{|S}$ is a compression of both G^S and G, so $\lambda_k(G^{|S}) \leq \lambda_k(G), \lambda_k(G^S)$ by the Cauchy interlacing theorem. Noticing $[A(G^S)]_{ij} \leq [A(G)]_{ij}$, $\lambda_1(G^S) = \sup_{\alpha\in S^{n-1}}\alpha' A(G^S)\alpha \leq \sup_{\alpha\in S^{n-1}}\sum\alpha_i\alpha_j[A(G)]_{ij} = \lambda_1(G)$.

Step 2: It is clear that $\operatorname{Tr}(A(G^S)^k) < \operatorname{Tr}(A^k)$, since we can partition

$$n^{-1}\sum_{i_1,..,i_k\in V^k} A_{i_1i_2}...A_{i_ki_1} = n^{-1}\sum_{\mathcal{A}} A_{i_1i_2}...A_{i_ki_1} + \frac{1}{n}\sum_{V^k\smallsetminus\mathcal{A}} A_{i_1i_2}...A_{i_ki_1}$$

with $\mathcal{A} = \{i_1,...,i_k : \forall t\in[k],\ i_t\vee i_{t+\epsilon(t)}\in S,\ \epsilon(t)\in\{-1,1\}\}$ and $\sum_{V^k\smallsetminus\mathcal{A}} A(G^S)_{i_1i_2}...A(G^S)_{i_ki_1} = 0$. Let $S_{k,\sigma}$ be the set of all k-sequences using elements from σ. Let η_j be the number of terms in the sum with j distinct nodes. Then notice $\frac{1}{n}\sum\eta_j = \mu^k$. For $A(G^{|S})$, we have

$$\begin{aligned}
\mathrm{E}\left[\mu^k(G^{|S})|G\right] &= m^{-1}\sum_{\sigma\in S}\sum_{i_1,...,i_k\in S_{k,\sigma}} A_{i_1i_2}...A_{i_ki_1}\mathrm{P}\left(\sigma\right)\\
&= m^{-1}\mathrm{P}\left(\sigma\right)\sum_{\sigma}\sum_{j=2}^{k}\left(\begin{array}{c} n-j\\ m-j\end{array}\right)\eta_j = m^{-1}\sum_{j=2}^{k}\frac{(m)_j}{(n)_j}\eta_j = n^{-1}\sum_{j=2}^{k}\frac{(m-1)_j}{(n-1)_j}\eta_j,
\end{aligned}$$

completing the proof. ■

Proof of Proposition 3.4. We use T to denote the row-stochastized adjacency matrix, $T_{ij} = A_{ij}/d_i$, instead of w, where d_i is the degree of node i to follow the literature (e.g., Jackson, 2008b). Let $\bar{T} = T(\widetilde{G})$.

Step 1: We show the argument for the case with G^S. The argument for $G^{|S}$ is similar, but omitted. Let

$$\bar{u} = M^0\left(\rho(T-\bar{T})y + \delta(T-\bar{T})x + \epsilon\right) = M^0 u,$$

where $M^0 = I_n - \iota\iota'/n$. The instrument is $\bar{Z} = [\iota, x, \bar{T}x, \bar{T}^2x]$. It suffices to show $\mathrm{E}[\bar{Z}'M^0u] \neq 0$. We can write

$$\mathrm{E}\left[\bar{Z}'\bar{u}|x,G\right] = \mathrm{E}\left[\left(\iota' M^0u, x'M^0u, x'\bar{T}'M^0u, x'\bar{T}^2M^0u\right)'\middle| x, G\right].$$

The first two components mechanically have expectation zero. By the reduced form representation in section 3.2, we can write y as a function of x and powers of T. The third component requires considering terms of the form $x'\mathrm{E}\left[\bar{T}'M^0\left(\bar{T}-T\right)|x,G\right]\delta x$. For generic x, this is zero if and only if $\mathrm{E}\left[\bar{T}'M^0\left(\bar{T}-T\right)|x,G\right]=0$. If we show $\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right)\neq\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)$, the preceding equation does not hold. We pass the expectation by linearity, use a cyclic permutation and write

$$\mathrm{E}\left[\mathrm{Tr}(\bar{T}'M^0\bar{T})\right]=\mathrm{E}\left[\mathrm{Tr}(M^0\bar{T}\bar{T}')\right]=\mathrm{E}\left[\mathrm{Tr}(\bar{T}\bar{T}')\right]-n^{-1}\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T})\right].$$

Let $\langle\cdot,\cdot\rangle_F$ be the Frobenius inner product, $\langle A,B\rangle_F=\mathrm{Tr}\left(AB'\right)$ and $\|\cdot\|_F$ the Frobenius norm, $\|A\|_F^2=\mathrm{Tr}(AA')$. In Lemmas D.7, D.8, D.9, and D.10 we compute the following four terms $\mathrm{E}\left[\|\bar{T}\|_F^2\right]$, $\mathrm{E}\left[\left\langle T,\bar{T}\right\rangle_F\right]$, $n^{-1}\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T})\right]$, and $n^{-1}\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'T)\right]$ which we then use to complete the argument. We find

$$\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)=(1-n^{-1})\{\|T\|_F^2+\sum_i\xi_2(d_i,\psi)\}-n^{-1}\xi_4(\vec{d},\psi),\ \mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right)$$
$$=(1-n^{-1})\{\|T\|_F^2+\sum_i\xi_1(d_i,\psi)\}-n^{-1}\xi_3(\vec{d},\psi).$$

In Lemma B.2 we show that $(n-1)\sum_i\left\{\xi_2(d_i,\psi)-\xi_1(d_i,\psi)\right\}-\left\{\xi_4(\vec{d},\psi)-\xi_3(\vec{d},\psi)\right\}\neq 0$ for all but finitely many $\psi\in(0,1)$, with an upper bound of $2\cdot\max_i d_i$ points, which completes the argument.

Step 2: We now show that the restriction of the set of observations in the second stage to $i\in S$ yields $\mathrm{E}\left[Z_{G^S}u_{G^S}\right]=0$. This follows from the fact that $\bar{T}y=Ty$ for all such $i\in S$, and therefore $\bar{u}=\epsilon$. The result follows from the fact that the instrument is correlated with Ty but orthogonal to ϵ, despite measurement error. ■

Lemma B.2. *Given a graph with non-degenerate coefficients above, for only a finite number of* $\psi\in(0,1)$ *can* $\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right)\neq\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)$.

Proof. Let $f\left(\psi,G\right):=(n-1)\sum_i\left\{\xi_2(d_i,\psi)-\xi_1(d_i,\psi)\right\}-\left\{\xi_4(\vec{d},\psi)-\xi_3(\vec{d},\psi)\right\}$. Showing $\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right)\neq\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)$ is equivalent to showing $f(\psi,G)\neq 0$. By the definitions of ξ_k, $f(\psi,G)$ is a rational function in ψ, with a numerator polynomial of degree bounded by $2\cdot\max_i d_i$ and coefficients given by G. As we assume that the graph does not yield coefficients so that the rational function is degenerate at zero, by the fundamental theorem of algebra, there are at most $2\cdot\max_i d_i$ roots of the numerator polynomial in ψ. This bounds the number of sampling rates $\psi\in(0,1)$ that would exactly satisfy the exclusion restriction for G. ■

Proof of Proposition 3.5.

Step 1: Let β^* solve $\bar{\rho}=\sum_d\frac{\beta^*\sigma(\beta^*)d}{1+\beta^*\sigma(\beta^*)d}\bar{\mathrm{P}}(d)$ where $\bar{\mathrm{P}}(d)$ is a sampled degree distribution. By (3.1) and that $\frac{\beta\sigma(\beta)d}{1+\beta\sigma(\beta)d}$ is strictly increasing in d when $\beta>0$, we have $\beta^*>\beta_0$ provided first order stochastic dominance of $\mathrm{P}(d)$ over $\bar{\mathrm{P}}(d)$. For G^S, for every d the count of nodes with at most degree d is weakly increasing under sampling; first order stochastic dominance follows. For $G^{|S}$ the argument is more delicate and relies on this being true in the limit. We use $\limsup_{R\to\infty}\sup_{r\le R}\sup_d |\mathrm{P}_{Rr}(d)-\mathrm{P}_{\infty r}(d)|=0$ which we did not need for the star subgraph, which implies $\limsup_{R\to\infty}\sup_{r\le R}\sup_d \left|\mathrm{P}^{|S}_{Rr}(d)-\mathrm{P}^{|S}_{\infty r}(d)\right|=0$. Let $F_{|S}(x)$ be the CDF for $\mathrm{P}^{|S}_\infty$ and $F(x)$ for P_∞.

$$\begin{aligned}
F_{|S}(x) &= \sum_{d\le x}\sum_{i\ge d}\mathrm{P}(i)\binom{i}{d}\psi^d(1-\psi)^{i-d} = \sum_{i=1}^{\infty}\mathrm{P}(i)\sum_{d\le i\wedge x}\binom{i}{d}\psi^d(1-\psi)^{i-d}\\
&= \sum_{i=1}^{x}\mathrm{P}(i)\cdot F_{Bin(i,\psi)}(i) + \sum_{i=x+1}^{\infty}\mathrm{P}(i)\cdot F_{Bin(i,\psi)}(x)\\
&= \sum_{i=1}^{x}\mathrm{P}(i) + \sum_{i=x+1}^{\infty}\mathrm{P}(i)\cdot F_{Bin(i,\psi)}(x) \ge \sum_{d\le x}\mathrm{P}(d) = F(x),
\end{aligned}$$

which confirms the stochastic dominance. The usual argument for GMM consistency shows $\operatorname{plim}\widehat{\beta}>\beta_0$ since in the limit β^* for every graph is greater than β_0, proving the result.

Step 2: By Jackson and Rogers (2007b), in graph r infection can spread only if $\beta_0>\zeta_{rR}^{-1}$. By arguments analogous to those in Lemma D.6,

$$\mathrm{E}[d(G^{|S})]/\mathrm{E}[d^2(G^{|S})] = \frac{\psi\mathrm{E}d}{\psi^2\mathrm{E}d^2+(1-\psi)\psi\mathrm{E}d}+o(1) = \zeta^{-1}+(1-\psi)\cdot\underbrace{\frac{1+\zeta^{-1}}{\psi\zeta+(1-\psi)}}_{\text{Positive}}+o(1) \text{ and}$$

$$\mathrm{E}[d(G^{S})]/\mathrm{E}[d^2(G^{S})] = \frac{\psi(2-\psi)\mathrm{E}d}{\psi\mathrm{E}d^2+\psi(1-\psi^2)\mathrm{E}d}+o(1) = \zeta^{-1}+(1-\psi)\underbrace{\left\{\frac{1-\zeta^{-1}(1+\psi)}{\zeta+(1-\psi^2)}\right\}}_{\text{Positive if }\zeta>(1+\psi).}+o(1).$$

The result follows since, by assumption on δ_{rR}, $\beta_0<\zeta^{-1}(\widetilde{G}_{rR})$ w.p.a.1 for every r, for $\widetilde{G}$ either $G^{|S}$ or G^S. ■

Appendix C. Proofs for Section 4

Proof of Lemma 4.1. For every network $a_n\cdot(\widehat{\theta}_r-\theta_{0r}) = -\left(\nabla_\theta\widehat{V}_{(r)}(\bar{\theta}_r)\right)^{-1}\cdot a_n\cdot\widehat{V}(\theta_{0r})$. By the Lipschitz condition 4 of the Lemma and Lemma E.2, $\sup_r\left\|\nabla_\theta\widehat{V}_{(r)}(\bar{\theta}_r)-\nabla_\theta\widehat{V}_{(r)}(\theta_{0r})\right\|=o_{\mathrm{P}}(1)$, which can be seen from

$$\sup_r\left\|\nabla_\theta\widehat{V}_{(r)}(\bar{\theta}_r)-\nabla_\theta\widehat{V}_{(r)}(\theta_{0r})\right\| \quad\le\quad \sup_r\|B_r\|\cdot\sup_r\left\|\widehat{\theta}_r-\theta_{0r}\right\| = o_{\mathrm{P}}(1).$$

By condition 5 of the Lemma $\sup_r\left\|a_n\cdot(\hat{\theta}_r-\theta_{0r})+\left(\nabla_\theta V_{(r)}(\theta_{0r})\right)^{-1}\cdot a_n\cdot\widehat{V}(\theta_{0r})\right\|=o_{\mathrm{P}}(1)$. By condition 3, we have that $a_n\cdot\sup_r\|\widehat{V}_{(r)}(\theta_r)\|=O_{\mathrm{P}}(R^{1/b})$, since $\mathrm{E}[\|\sup_r a_n\cdot\widehat{V}_{(r)}(\theta_{0r})\|^b]^{1/b}\le$

$R^{1/b}\sup_r \|a_n \cdot \widehat{V}_{(r)}(\theta_{0r})\|_{\ell_b}$ and therefore $a_n \cdot \sup_r \left\|\widehat{\theta}_r - \theta_{0r}\right\| = O_{\mathrm{P}}(R^{1/b})$. If instead of b moments we assume all moments exist, then $\sqrt{\log(R+1)}$ replaces $R^{1/b}$ by a standard Orlicz inequality (e.g., Van der Vaart and Wellner, 1996). ■

C.1. **Proof of Theorem 4.1.** The argument is entirely standard and follows by expanding around $\widehat{\theta}$ uniformly and then checking that conditional expectations preserve asymptotic normality. We separate the OLS and GMM cases.

Proof of Theorem 4.1.1. Consistency is clear so we directly check normality. Let $u_{ir} = \epsilon_{ir} + (w_{ir}(G_r) - \mathcal{E}_{ir}(x_r;\widehat{\theta}_r))\beta_0$ and $H_R(\widehat{\theta}) := (nR)^{-1}\mathcal{E}(x;\widehat{\theta})'\mathcal{E}(x;\widehat{\theta})$. As usual $\sqrt{nR}\left(\widehat{\beta} - \beta_0\right) = H_R(\widehat{\theta})^{-1} \cdot (nR)^{-1/2}\mathcal{E}(x;\widehat{\theta})'u$.

Step 1: To show $H_R(\widehat{\theta}) - H_{R,0}(\theta_0) = o_{\mathrm{P}}(1)$, where $H_{R,0}(\theta_0) := \mathbb{E}_{nR}\left[\mathrm{E}[\mathcal{E}_{ir}(x_r;\theta_{0r})\mathcal{E}_{ir}(x_r;\theta_{0r})']\right]$. By a term-by-term expansion,

$$H_R(\widehat{\theta}) = H_R(\theta_0) + \mathbb{E}_{nR}\left[\dot{\Psi}_{ir}(\bar{\theta}_r)\left(I_p \otimes (\widehat{\theta}_r - \theta_{0r})\right)\right]$$

where $\dot{\Psi}_{ir}(\bar{\theta}_r) = \frac{\partial}{\partial\theta'}\left\{\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\mathcal{E}_{ir}(x_r;\bar{\theta}_r)'\right\}$ is a $p \times pk$ matrix of derivatives w.r.t. θ_{rk}, which exists by Assumption A.3. By Lemma E.4 the second term $o_{\mathrm{P}}(1)$ since $\left\|\dot{\Psi}_{ir}(\bar{\theta}_r)\right\| \lesssim \left\|\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \left\|\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \lesssim L_{i,r}(x_r)^2$, which by Assumption A.4.2 follows from

$$\left\|\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \leq L_{i,r}(x_r) \text{ and } \left\|\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \leq \sqrt{\mathrm{E}\left[\|w_{ir}\|^2 \,|x_r;\bar{\theta}_r\right]}\cdot\sqrt{\mathrm{E}\left[\left\|\mathcal{I}_{w_{ir}|x_r}(\bar{\theta}_r)\right\| |x_r;\bar{\theta}_r\right]} \leq L_{i,r}(x_r).$$

That $H_R(\theta_0) - H_{R,0}(\theta_0) = o_{\mathrm{P}}(1)$ follows from Lemma E.3.

Step 2: We show $(nR)^{-1/2}\mathcal{E}(x;\widehat{\theta})'u \rightsquigarrow \mathcal{N}(0,V)$. Let $f_{ir}(\theta_r;\beta) := \mathcal{E}_{ir}(x_r;\theta_r)(w_{ir}(G_r) - \mathcal{E}_{ir}(x_r;\theta_r))'\beta$. Then

$$\mathcal{E}(x;\widehat{\theta})'u/\sqrt{nR} = f(\theta_0;\beta_0)/\sqrt{nR} + \mathcal{E}(x;\theta_0)'\epsilon/\sqrt{nR} + \sqrt{nR}\mathbb{E}_{nR}\left[\dot{\Phi}_{ir}(\bar{\theta}_r)\left(I_p \otimes (\widehat{\theta}_r - \theta_{0r})\right)\right]$$

where $\dot{\Phi}_{ir}(\bar{\theta}_r) = \frac{\partial}{\partial\theta_r}\left\{f_{ir}(\bar{\theta}_r;\beta_0) + \mathcal{E}_{ir}(x_r;\bar{\theta}_r)\epsilon_{ir}\right\}$.

Clearly $\mathrm{E}\left[f_{ir}(\theta_{0r},\beta_0)\right] = 0$ and $\mathrm{E}\left[\mathcal{E}_{ir}(x_r;\theta_{0r})\epsilon_{ir}\right] = 0$.[52] Let $g_{ir}(\theta_{0r};\beta_0) := f_{ir}(\theta_{0r};\beta_0) + \mathcal{E}_{ir}(x_r;\theta_{0r})\epsilon_{ir}$. That $\sqrt{nR}\mathbb{E}_{n,R}g_{ir}(\theta_{0r},\beta_0) \rightsquigarrow \mathcal{N}(0,V)$, where $V = \lim \mathbb{E}_R\mathrm{var}\left(\sqrt{n}\mathbb{E}_n g_{ir}(\theta_{0r},\beta_0)\right)$, follows from Lemma E.4 which can be applied by Assumption A.4.3.

Finally, the $\dot{\Phi}_{ir}(\bar{\theta}_r)$ term is controlled by Lemma E.1 using $\left\|\frac{\partial}{\partial\theta'}f_{ir}(\bar{\theta}_r;\beta_0)\right\| \lesssim_{\mathrm{P}} L_{i,r}(x_r)^2$, which follows from

$$\begin{aligned}\left\|\frac{\partial}{\partial\theta'}f_{ir}(\bar{\theta}_r;\beta_0)\right\| &\leq \sup_{\beta\in\mathcal{B}}\|\beta\| \cdot \left\|\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \left\|w_{ir}(G_r) - \frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \\ &\lesssim L_{i,r}(x_r)\|w_{ir}(G_r)\| + L_{i,r}(x_r)^2 \lesssim_{\mathrm{P}} L_{i,r}(x_r)^2\end{aligned}$$

as $\|w_{ir}(G_r)\| \lesssim_{\mathrm{P}} \mathrm{E}\left[\|w_{ir}(G_r)\|\right] \leq \mathrm{E}\left[L_{i,r}(x_r)\right]$. ■

[52]The former by iterated expectations and the latter by assumption on ϵ.

Next, we turn to GMM. The argument is conceptually similar and depends on uniform expansions. The main difference is control of score functions. Let us define the conditional score[53] with respect to β and the conditional score with respect to θ_r as $S_{m_{ir}|x_r}(\beta;\theta_r) := \frac{\partial}{\partial\beta'}\log f(m_{ir}|x_r;\beta,\theta_r)$ and $S_{m_{ir}|x_r}(\theta_r;\beta) = \frac{\partial}{\partial\theta_r'}\log f(m_{ir}|x_r;\beta,\theta_r)$. The corresponding information matrices are

$$\mathcal{I}_{m_{ir}|x_r}(\beta;\theta_r) = \mathrm{E}\left[S_{m_{ir}|x_r}(\beta;\theta_r)S_{m_{ir}|x_r}(\beta;\theta_r)'|x_r;\beta,\theta_r\right]$$

and

$$\mathcal{I}_{m_{ir}|x_r}(\theta_r;\beta) = \mathrm{E}\left[S_{m_{ir}|x_r}(\theta_r;\beta)S_{m_{ir}|x_r}(\theta_r;\beta)'|x_r;\beta,\theta_r\right].$$

Notice that $\left\|\mathcal{I}_{m_{ir}|x_r}(\beta;\theta_r)\right\| \leq \left\|\mathcal{I}_{m_{ir}|x_r}(\beta,\theta_r)\right\|$ and $\left\|\mathcal{I}_{m_{ir}|x_r}(\theta_r;\beta)\right\| \leq \left\|\mathcal{I}_{m_{ir}|x_r}(\beta,\theta_r)\right\|$, since each is a projection of the larger information matrix, by Cauchy's interlacing theorem. We also use a shorthand $\mathcal{E}_{ir}(\beta,\theta_r)$ for $\mathcal{E}_{ir}(x_r;\beta,\theta_r)$.

Lemma C.1. *Under Assumptions A.3 and A.5,* $\widehat{\beta}_{\mathrm{gmm}} \xrightarrow{\mathrm{P}} \beta_0$.

Proof. In four steps we check conditions 1-4 of Andrews (1994), Theorem A-1.

Step 1: The first part is clear by Assumption A.5.2, since $\mathrm{E}\left[m(y_{ir},w_{ir};\beta)|x_r;\theta_r,\beta\right] = \mathcal{E}_{ir}(\beta,\theta_r)$. That $\mathcal{E}_{ir}(\beta,\theta_r)$ satisfies a uniform law of large numbers,

$$\sup_{(\beta,\theta)\in\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r}\left\|\mathbb{E}_{nR}\left\{\mathcal{E}_{ir}(\beta,\theta_r) - \mathrm{E}\mathcal{E}_{ir}(\beta,\theta_r)\right\}\right\| = o_{\mathrm{P}}(1),$$

follows from a pointwise convergence, which is clear, and stochastic equicontinuity. Stochastic equicontinuity follows from a Lipschitz condition. Define the following.

$$\epsilon^1_{i,r}(\bar\beta,\beta,\theta_r') := \mathrm{E}\left[\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir}(G_r);\bar\beta)|z_r,A^o_r,y_r;\theta_r,\beta\right],$$

$$\epsilon^2_{i,r}(\beta',\tilde\beta,\theta_r') := \mathrm{E}\left[m(y_{ir},w_{ir}(G_r);\beta')\cdot S_{m_{ir}|x_r}(\tilde\beta;\theta_r')'|z_r,A^{obs}_r,y_r;\theta_r,\tilde\beta\right],$$

$$\text{and } \epsilon^3_{i,r}(\beta',\bar\theta_r,\theta_r) := \mathrm{E}\left[m(y_{ir},w_{ir}(G_r);\beta')\cdot S_{m_{ir}|x_r}(\bar\theta_r;\beta')'|z_r,A^{obs}_r,y_r;\theta_r,\beta'\right].$$

By a Taylor expansion it follows that

$$\left\|\mathrm{E}\left[m(y_{ir},w_{ir}(G_r);\beta')|z_r,A^{obs}_r,y_r;\theta_r',\beta'\right] - \mathrm{E}\left[m(y_{ir},w_{ir}(G_r);\beta)|z_r,A^{obs}_r,y_r;\theta_r,\beta\right]\right\|$$

is bounded by $\left\|\epsilon^1_{i,r}(\bar\beta,\beta,\theta_r') + \epsilon^2_{i,r}(\beta',\tilde\beta,\theta_r')\right\|\|\beta'-\beta\| + \left\|\epsilon^3_{i,r}(\bar\beta,\theta_r')\right\|\|\theta_r'-\theta_r\|$. Then we have by Assumption A.5.3,

$$\left\|\epsilon^1_{i,r}(\bar\beta,\beta,\theta_r')\right\| \leq \mathrm{E}\left[\left\|\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir}(G_r);\bar\beta)\right\||x_r;\theta_r',\beta\right] \leq L_{i,r}(x_r),$$

[53]The right-hand side of the expression abuses notation. Given that graphs occupy a discrete space, with continuously distributed disturbances in the model we may formally have to decompose the measure into its absolutely continuous part and its pure point part (e.g., Lebesgue decomposition theorem). We rule out the singular part by assumption.

$$\left\|\epsilon^2_{i,r}(\beta',\tilde\beta,\theta'_r)\right\| \le \sqrt{\mathrm{E}\left[\|m_{ir}(\beta')\|^2\,|x_r;\theta'_r,\tilde\beta\right]}\cdot\sqrt{\left\|\mathcal{I}_{m_{ir}|x_r}(\tilde\beta;\theta'_r)\right\|} \le L_{i,r}(x_r), \text{ and}$$

$$\left\|\epsilon^3_{i,r}(\beta',\bar\theta_r,\theta_r)\right\| \le \sqrt{\mathrm{E}\left[\|m_{ir}(\beta')\|^2\,|x_r;\theta'_r,\tilde\beta\right]}\cdot\sqrt{\left\|\mathcal{I}_{m_{ir}|x_r}(\bar\theta_r;\beta')\right\|} \le L_{i,r}(x_r).$$

As such, with $d_{\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r}((\beta,\theta),(\beta',\theta')) := \|\beta-\beta'\|\vee\sup_{r\in\mathbb{N}}\|\theta_r-\theta'_r\|$,

$$\|\mathbb{E}_{n,R}\left\{\mathcal{E}_{ir}(\beta,\theta_r)-\mathcal{E}_{ir}(\beta',\theta'_r)\right\}\| \le \mathbb{E}_{n,R}L_{i,r}\cdot\sup_{r\le R}\left(\|\beta'-\beta\|+\|\theta'_r-\theta_r\|\right) \le \mathbb{E}_{n,R}L_{i,r}\cdot d_{\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r}((\beta,\theta),(\beta',\theta'))$$

Since $\sup_{R\ge1}\mathbb{E}_{n,R}[\mathrm{E}L_{i,r}] < \infty$, by Andrews (1992) Lemma 2 it follows that stochastic equicontinuity holds and that $\mathbb{E}_{n,R}\mathrm{E}\mathcal{E}_{ir}(\beta,\theta_r)$ is uniformly continuous in $(\beta,\theta)\in\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r$.

Step 2: The second and third part of the second condition are clear: by Assumption A.5.1 $\widehat{W}-W = o_{\mathrm{P}}(1)$ and by Assumption A.3 $\left(\widehat\theta_1,...,\widehat\theta_R\right)\in\prod_{r=1}^R\Theta_r$. We need only show $\sup_{\beta\in\mathcal{B}}\left\|\lim_{R\to\infty}\mathbb{E}_{n,R}\mathrm{E}\mathcal{E}_{ir}(\beta,\widehat\theta_r)-\lim_{R\to\infty}\mathbb{E}_{n,R}\mathrm{E}\mathcal{E}_{ir}(\beta,\theta_{0r})\right\| = o_{\mathrm{P}}(1)$ for $\theta_0\in\prod_{r\in\mathbb{N}}\Theta_r$. This follows from a Taylor expansion and the fact that

$$\sup_{\beta\in\mathcal{B}}\lim_{R\to\infty}\mathbb{E}_{n,R}\left\|\mathrm{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar\theta_r)\cdot\left(I\otimes(\widehat\theta_r-\theta_{0r})\right)\right\| \le \limsup_{R\to\infty}\mathbb{E}_{n,R}\mathrm{E}L_{i,r}\cdot\sup_r\left\|\widehat\theta_r-\theta_{0r}\right\| = o_{\mathrm{P}}(1)$$

where $\sup_r\left\|\hat\theta_r-\theta_{0r}\right\| = o_{\mathrm{P}}(1)$ by Assumption A.3.3 and $\limsup_{R\to\infty}\mathbb{E}_{R,R}\mathrm{E}L_{i,r} < \infty$ by A.5.3. This, in turn, results from the fact that $\left\|\mathrm{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar\theta_r)\right\| \le \mathrm{E}L_{i,r}$. To see this, note $\left\|\mathrm{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar\theta_r)\right\| \le \mathrm{E}\left[\left\|\mathrm{E}\left[m_{ir}(\beta)S_{m_{ir}|x_r}(\beta;\bar\theta_r)|x_r;\beta,\bar\theta_r\right]\right\|\right]$. By Assumptions A.5.3 and A.5.4, $\left\|\mathrm{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar\theta_r)\right\| \le \mathrm{E}\left[\sqrt{\left(\mathrm{E}\left[\|M_{i,r}(X_{ir})\|^2\,|x_r;\beta,\bar\theta_r\right]\right)}\cdot\sqrt{\left\|\mathcal{I}_{m_{ir}|x_r}(\beta;\bar\theta_r)\right\|}\right] \le \mathrm{E}L_{i,r}(x_r)$.

Step 3: This follows from $\mathrm{E}\|\mathcal{E}_{ir}(\beta,\theta_r)\| \le \mathrm{E}L_{i,r}$ and $\limsup_{n\to\infty}\mathbb{E}_{n,R}\mathrm{E}L_{i,r} < \infty$.

Step 4: By iterated expectations $g_n(\beta) = \mathbb{E}_{n,R}\mathrm{E}\left[\mathrm{E}\left[m(X_{ir};\beta)|x_r;\theta_{0r},\beta_0\right]\right] = \mathbb{E}_{n,R}\mathrm{E}\eta_{ir}(\beta,\theta_{0r},\beta_0)$ where $\eta_{ir}(\beta,\theta_r,\beta) := \mathcal{E}_{ir}(\beta,\theta_r)$. By the identification condition, Assumption A.5.1

$$W\lim_{R\to\infty}\mathbb{E}_{n,R}\mathrm{E}\eta_{ir}(\beta,\theta_{0r},\beta_0)\neq0 \text{ and } W\lim_{R\to\infty}\mathbb{E}_{n,R}\mathrm{E}\mathcal{E}_{ir}(\beta,\theta_{0r})\neq0$$

for $\beta\neq\beta_0$, but $W\lim_{R\to\infty}\mathbb{E}_{n,R}\mathrm{E}\mathcal{E}_{ir}(\beta_0,\theta_{0r}) = W\lim_{R\to\infty}\mathbb{E}_{n,R}\mathrm{E}\eta_{ir}(\beta_0,\theta_{0r},\beta_0) = 0$. By positive semi-definiteness of W, letting $K'K = W$, observe that $0\neq Wf(\beta) = K'Kf(\beta)$ implies $Kf(\beta)\neq0$. It follows that $Q(\beta,\theta_0,W) > Q(\beta_0,\theta_0,W)$ for any $\beta\neq\beta_0$. ■

Proof of Theorem 4.1.2.

Step 1: The estimator satisfies for $\widehat\gamma_{n,R}(\widehat\beta;\widehat\theta) := \mathbb{E}_{n,R}\left[\mathcal{E}_{ir}(\widehat\beta,\widehat\theta_r)\right]$,

$$\left[\frac{\partial}{\partial\beta'}\widehat\gamma_{n,R}(\widehat\beta;\widehat\theta)\right]'\widehat{W}\sqrt{nR}\widehat\gamma_{n,R}(\widehat\beta;\widehat\theta) = o_{\mathrm{P}}(1).$$

A term-by-term expansion yields $\sqrt{nR}\widehat\gamma_{n,R}(\widehat\beta;\widehat\theta) = \sqrt{nR}\widehat\gamma_{n,R}(\widehat\beta;\theta_0^R)+\sqrt{nR}\mathbb{E}_{n,R}\left[\dot\Psi_{ir}(\bar\theta_r;\widehat\beta)\left(I_q\otimes(\widehat\theta_r-\theta_{0r})\right)\right]$, where $\dot\Psi_{ir}(\bar\theta_r;\widehat\beta) = \epsilon^3_{i,r}(\widehat\beta,\bar\theta_r,\bar\theta_r)$ has been controlled in the preceding lemma. Similarly

$\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\widehat{\beta};\widehat{\theta}) = \frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\widehat{\beta};\theta_0^R) + \mathbb{E}_{n,R}\left[\dot{\Phi}_{ir}(\bar{\theta}_r;\widehat{\beta})\left(I_q \otimes (\widehat{\theta}_r - \theta_{0r})\right)\right],$

$$\dot{\Phi}_{ir}(\bar{\theta}_r;\widehat{\beta}) = \mathrm{E}\left[\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir};\widehat{\beta})S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})|A_r^o,y_r,z_r;\bar{\theta}_r,\widehat{\beta}\right] + \mathrm{E}\left[m(y_{ir},w_{ir};\widehat{\beta})S_{m_{ir}|x_r}(\widehat{\beta};\bar{\theta}_r)S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})|A_r^o,y_r,z_r;\bar{\theta}_r,\widehat{\beta}\right].$$

We can bound the first term by $(\mathrm{E}[\|\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir};\widehat{\beta})\|^2|x_r;\bar{\theta}_r,\widehat{\beta}]\cdot\mathrm{E}[\|S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})\|^2|x_r;\bar{\theta}_r,\widehat{\beta}])^{1/2}$ and the second by

$$\sqrt{\mathrm{E}\left[\left\|\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir};\widehat{\beta})\right\|^2|x_r;\bar{\theta}_r,\widehat{\beta}\right]}\sqrt{\sqrt{\mathrm{E}\left[\left\|S_{m_{ir}|x_r}(\widehat{\beta};\bar{\theta}_r)\right\|^4|x_r;\bar{\theta}_r,\widehat{\beta}\right]}\cdot\sqrt{\mathrm{E}\left[\left\|S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})\right\|^4|x_r;\bar{\theta}_r,\widehat{\beta}\right]}}$$

and therfore $\left\|\dot{\Phi}_{ir}(\bar{\theta}_r;\widehat{\beta})\right\| \le 2L_{i,r}(x_r)$. By Lemma E.1, $\left[\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\widehat{\beta};\theta_0^R)\right]'\widehat{W}\sqrt{nR}\widehat{\gamma}_{n,R}(\widehat{\beta};\theta_0^R) = o_{\mathrm{P}}(1)$.

Step 2: A term-by-term expansion for the jth term yields $\sqrt{nR}\widehat{\gamma}^j_{n,R}(\widehat{\beta};\theta_0^R) = \sqrt{nR}\widehat{\gamma}^j_{n,R}(\beta_0;\theta_0^R) + \sqrt{nR}\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R)\left(\widehat{\beta}-\beta_0\right)$, so it remains to be seen that $\left\|\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R) - M\right\| = o_{\mathrm{P}}(1)$, where $M = \lim_{n\to\infty}\mathbb{E}_{nR}\left[\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right]$, and that $\sqrt{nR}\widehat{\gamma}_{n,R}(\beta_0;\theta_0^R) \rightsquigarrow \mathcal{N}(0,\Omega)$, where $\Omega := \lim_{n\to\infty}\mathbb{E}_R\left[\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right]$. For the derivative,

$$\begin{aligned}\left\|\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R) - M\right\| \le{}& \left\|\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R) - \mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r})\right\| \\ &+ \left\|\mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r}) - \lim_{n\to\infty}\mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r})\right\| \\ &+ \left\|\lim_{n\to\infty}\mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r}) - M\right\|.\end{aligned}$$

The first term is $o_{\mathrm{P}}(1)$ by a uniform law of large numbers for the derivative which exists by Assumption A.5.4 and the second term is $o_{\mathrm{P}}(1)$ by as the limit exists uniformly over $\mathcal{B}$ by Assumption A.5.2. The final term is $o_{\mathrm{P}}(1)$ as $\tilde{\beta}-\beta_0 = o_{\mathrm{P}}(1)$ by continuity.

To show $\sqrt{nR}\widehat{\gamma}_{n,R}(\beta_0;\theta_0^R) \rightsquigarrow \mathcal{N}(0,\Omega)$, by Assumption A.5.5 we may apply Lemma E.4 with $m(X_{ir};\beta_0) = Y_{ir}$, $\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r}) = Z_{ir}$, and $x_r = X_r$ in the notation of the lemma. It is clear that $\Omega := \lim_{n\to\infty}\mathbb{E}_R\left[\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right]$, exists under these conditions since for every r, $\lambda_{\max}\left(\mathrm{var}\left(\sqrt{n}\mathbb{E}_n m(X_{ir};\beta_0)\right)\right) \ge \lambda_{\max}\left(\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right)$. ■

C.2. **Example 1: Conditional Edge Independent Models.**

Proof of Proposition 4.1. We check the conditions of Lemma 4.1. The first condition is clear by definition and Assumption A.7. Condition 2 is shown in Lemma E.6. We show directly $\sup_r\left||\Xi|^{-1/2}\,\mathbb{E}_\Xi[v(X_{rs};\theta_{0r})]\right| = O_{\mathrm{P}}(R^{1/b})$ in Lemma E.5, which is what condition 3 is used for in the proof of Lemma 4.1. The Lipschitz condition, condition 4, follows from a secondary expansion and the assumption of the existence of an envelope function. That is,

$$|\Xi|^{-1}\sum_{s\in\Xi}\left\{\nabla_\theta v(X_{rs};\theta_r^*) - \nabla_\theta v(X_{rs};\theta_{0r})\right\} = |\Xi|^{-1}\sum_{s\in\Xi}F_{rs}\left(I\otimes(\theta_r^* - \theta_{0r})\right)$$

where F_{rs} is a conformable matrix of derivatives evaluated at an intermediate value and $F_{rs} \leq B$, which can be done by Assumption A.7.2. The Hessian condition, condition 5, hold by the envelope condition and Chebyshev's inequality. ∎

C.3. **Example 2: Links and Triangles SUGM.**

Proof of Proposition 4.2. We sketch the proof here for brevity, checking the conditions of Lemma 4.1. The argument is similar to the proof of Proposition 4.1. Condition 1 requires standard conditions for consistency: a unique maximum at θ_{0r}, compactness which is assumed, smoothness of derivatives (the moment function is analytic in parameters so this holds), and uniform law of large numbers, all of which are shown in Chandrasekhar and Jackson (2016), proof of Proposition 4 (in particular in proofs of Proposition D.1 which shows identification and D.2 which shows consistency). Condition 4 requires a Lipschitz condition, which is shown in Chandrasekhar and Jackson (2016), proof of Proposition D.2. Condition 3(i) follows from the the proof of asymptotic normality in Chandrasekhar and Jackson (2016) in Lemma D.4. Condition 3(ii) follows from an argument identical to that of Lemma E.5. Since the (appropriately normalized) moment function is asymptotically normally distributed, and because the summand is just a binary variable indicating presence or absence of a subgraph, then certainly any bth moment exists (all moments exist), so the same argument of Lemma E.5 applies. Condition 5 follows from the fact that the moment summand has an envelope with arbitrary moments and Chebyshev's inequality.

The main item to check is condition 2: that the objective functions uniformly converge. The argument is nearly identical to that in the proof of Lemma E.6. Where we then use Lemma E.7 to control the final term, we can replace it with the computations below.

Here it is useful to write the GMM objective function, call it $\widehat{Q}$ in terms of the moment functions, call them $\widehat{M}$, since we need to only control the moment functions, as is clear by adding and subtracting:

$$\begin{aligned}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| &= \left|\widehat{M}_{(r)}(\theta)' \widehat{M}_{(r)}(\theta) - M_{(r)}(\theta)' M_{(r)}(\theta)\right| \\ &\leq \left\|\widehat{M}_{(r)}(\theta)'\right\| \left\|\widehat{M}_{(r)}(\theta) - M_{(r)}(\theta)\right\| + \left\|\widehat{M}_{(r)}(\theta)' - M_{(r)}(\theta)'\right\| \left\|M_{(r)}(\theta)\right\|\end{aligned}$$

where

$$\widehat{M}_{(r)}(\theta) = \left(\frac{n^{h_L}}{\binom{n}{2}} \sum_{i<j} \{A_{ij} - \mathrm{E}_\theta A_{ij}\}, \frac{n^{h_T}}{\binom{n}{3}} \sum_{i<j<k} \{A_{ij}A_{ik}A_{jk} - \mathrm{E}_\theta A_{ij}A_{ik}A_{jk}\}\right)'.$$

To simplify notation let $X_{ij} = A_{ij} - \mathrm{E}A_{ij}$ and $X_{ijk} = A_{ij}A_{ik}A_{jk} - \mathrm{E}A_{ij}A_{ik}A_{jk}$.

The argument is identical to that in the proof of Lemma E.6, but in the final step instead of using the envelope B we can directly use the moment summands themselves. There we had to appeal to the random field that generated the covariates and show that under sensible assumptions, sets that are far apart are near independent.

In this case we can rely on computations in Chandrasekhar and Jackson (2016) that directly show under reasonable assumptions the covariance terms are small relative to the variance terms. By the proof of Lemma D.4 in Chandrasekhar and Jackson (2016), with S_L, the share of links, observe that (applying Corollary D.1), we can write

$$\mathrm{P}\left(\left|\frac{1}{\binom{n}{2}}\sum_{i<j}X_{ij}\right|>\eta n^{-h_L}\right)\leq\frac{\mathrm{var}\left(\sum X_{ij}\right)}{\binom{n}{2}^2\eta^2}n^{2h_L}\leq n^{2h_L}\frac{\binom{n}{2}\mathrm{var}\left(X_{ij}\right)+o\left(\binom{n}{2}\mathrm{var}\left(X_{ij}\right)\right)}{\binom{n}{2}^2\eta^2}$$
$$\lesssim n^{2h_L}\frac{\mathrm{var}\left(X_{ij}\right)}{\binom{n}{2}}=n^{2h_L}\frac{p_{0L}\left(1-p_{0L}\right)}{\binom{n}{2}}\lesssim\Theta\left(n^{h_L-2}\right).$$

Similarly, turning to S_T, the share of triangles,

$$\mathrm{P}\left(\left|\frac{1}{\binom{n}{3}}\sum_{i<j<k}X_{ijk}\right|>\eta n^{-h_T}\right)\lesssim n^{2h_T}\frac{\mathrm{var}\left(X_{ijk}\right)}{\binom{n}{3}}=n^{2h_T}\frac{p_{0T}\left(1-p_{0T}\right)}{\binom{n}{3}}\lesssim\Theta\left(n^{h_T-3}\right).$$

Then by a union bound, we can bound the order of the maximum:

$$\mathrm{P}\left(\max_r\left|\frac{1}{\binom{n}{2}}\sum_{i<j}X_{ij,r}\right|>\eta n^{-h_L}\right)\lesssim R\cdot\Theta\left(n^{h_L-2}\right)$$

and similarly

$$\mathrm{P}\left(\max_r\left|\frac{1}{\binom{n}{3}}\sum_{i<j<k}X_{ijk}\right|>\eta n^{-h_T}\right)\lesssim R\cdot\Theta\left(n^{h_T-3}\right).$$

Then for $R=n^k$, under our assumption that

$$k<\max\left\{2-h_L,3-h_T\right\},$$

we have uniform convergence of the objective function. ■

C.4. **Example 3: Groups Model.** The proof of Proposition 4.3 is a corollary to the following lemma. We need to replicate the arguments from section C.1 replacing Lemma E.1 by Lemma C.2. The model satisfies Assumption (2) below by Chatterjee et al. (2010).

LEMMA C.2. *Assume (1) for all $\theta_r\in\Theta_r$, $\zeta_r(\bar{\theta}_r)$ are $q\times k_nq$ matrix (or vector) valued functions, $\sup_r\|\zeta_r(\theta_r)\|\leq B_r$ with $\operatorname{plim}\mathbb{E}_{nR}[B_{ir}]<\infty$, (2) for each r, $\sup_{i\leq k_n}\left\|\widehat{\theta}_{ir}-\theta_{0ir}\right\|\leq C(L)\sqrt{\frac{\log n}{n}}$ with probability at least $1-cn^{-2}$, where L,c depend only on $\Theta_r\subset\mathbb{R}^{k_n}$ and ψ, (3) $\Theta_r\subset[a,b]^{k_n}$ for all r, and (4) $R=O(n^\gamma)$ with $\gamma<1$. Then*

$$R^{1/2}\left\|\mathbb{E}_R\zeta_r(\bar{\theta}_r)(I_q\otimes(\widehat{\theta}_r-\theta_{0r}))\right\|_2\lesssim\sqrt{\frac{R\cdot\log n}{n}}$$

with probability approaching one. Specifically the probability is at least $\left(1-cn^{-2}\right)^{Rn}$.

Proof of Lemma C.2.

Step 1: This follows from $\|\cdot\|_2 \leq \|\cdot\|_1$ and

$$\begin{aligned} R^{1/2}\left\|\mathbb{E}_R\zeta_r(\bar\theta_r)(I_q\otimes(\widehat\theta_r-\theta_{0r}))\right\|_1 &\leq R^{1/2}\mathbb{E}_R\left[\left\|\zeta_r(\bar\theta_r)\right\|_1\left\|(I_q\otimes(\widehat\theta_r-\theta_{0r}))\right\|_\infty\right] \\ &\leq R^{1/2}\mathbb{E}_R\left[\left\|\zeta_r(\bar\theta_r)\right\|_1\left\|\widehat\theta_r-\theta_{0r}\right\|_\infty\right] \\ &\leq \mathbb{E}_R\left\|\zeta_r(\bar\theta_r)\right\|_1\cdot\sqrt{R}\sup_{r\leq R}\left\|\widehat\theta_r-\theta_{0r}\right\|_\infty . \end{aligned}$$

We have that $\frac{1}{\sqrt{q}}\left\|\zeta_r(\bar\theta_r)\right\|_1 \leq \left\|\zeta_r(\bar\theta_r)\right\|_2 \leq B_r$ and therefore

$$R^{1/2}\left\|\mathbb{E}_R\zeta_r(\bar\theta_r)(I_q\otimes(\widehat\theta_r-\theta_{0r}))\right\|_1 \lesssim \sqrt{R}\sup_{r\leq R}\left\|\widehat\theta_r-\theta_{0r}\right\|_\infty + o_{\mathrm{P}}(1)$$

by assumption on the envelope functions. It suffices to control $\sup_{r\leq R}\left\|\widehat\theta_r-\theta_{0r}\right\|_\infty$.

Step 2: Observe that $\sup_{r\leq R}\left\|\widehat\theta_r-\theta_{0r}\right\|_\infty \leq C(L)\sqrt{\frac{\log n}{n}}$ with probability at least $(1-cn^{-2})^{R_n}$, since the bound holds for each of the R (independent) terms with probability at least $1-cn^{-2}$. This approaches one for $\gamma<1$. ■

Online Appendix: Not for Publication

Appendix D. Supplementary Proofs for Section 3

In what follows, we let $t(G) := \sum_{i<j<k} A_{ij}A_{jk}$ denote the number of two-stars ($A_{ij}A_{jk} = 1$), and $p(G) := \sum_{i<j<k<l} A_{ij}A_{kl}$ denote the number of disjoint pairs in the graph.

Lemma D.1. *Under uniform random sampling of m out of n nodes, with $m/n \to \psi$ and as $n \to \infty$,*

$\mathrm{E}[d(G^{|S})|G] = (\psi + \Theta(n^{-1}))d(G)$ *and* $\mathrm{var}(d(G^{|S})|G) = \Theta(n^{-1}d(G)) + \Theta(n^{-1}t(G)) + \Theta(n^{-2}p(G))$.

Proof. Step 1: Let $\mathcal{S}$ be the set of all combinations of m vertices with σ a generic element. Then

$$\mathrm{E}[d(G^{|S})|G] \quad = \quad \sum_{\sigma\in\mathcal{S}} d(\sigma)\mathrm{P}\left(\sigma\right) = \binom{n}{m}^{-1} \sum_{\sigma\in\mathcal{S}} m^{-1} \sum_{i\wedge j\in\sigma} A_{ij}.$$

Each pair ij appears $|\{\sigma\in\mathcal{S} : i\wedge j\in\sigma\}| = \binom{n-2}{m-2}$ times, it follows that $\mathrm{E}[d(G^{|S})|G] = \frac{m-1}{n-1}\cdot\frac{1}{n}\sum_{i,j} A_{ij} = (\psi + \Theta(n^{-1}))d(G)$.

Step 2: Use $|\{\sigma\in\mathcal{S} : i_1\wedge ... \wedge i_k \in\sigma\}| = \binom{n-k}{m-k}$ and let $\epsilon_1 := \sum A_{ij}$, $\epsilon_2 := \sum\{A_{ij}A_{kl} : i\vee j\in\{k,l\}, \neg i\wedge j\in\{k,l\}\ \}$, and $\epsilon_3 := \sum\{A_{ij}A_{kl} :\ \ i\notin\{j,k,l\}, j\notin\{k,l\}, k\neq l\ \}$. Then $\epsilon_1 = 2|E|$, $\epsilon_2 = 8t(G)$, $\epsilon_3 = 8p(G)$, and

$$\mathrm{E}[d(G^{|S})^2|G] = \frac{1}{m^2}\frac{m!}{(n)_m}\sum_{j=1}^{3}\left(\begin{array}{c} n-(j+1) \\ m-(j+1) \end{array}\right)\epsilon_j.$$

Some algebra yields $\mathrm{E}[d(G^{|S})^2|G] = \frac{1}{m}\frac{(m-1)}{(n-1)}\frac{2|E|}{n} + \frac{1}{m}\frac{(m-1)_2}{(n-1)_2}\frac{8t(G)}{n} + \frac{1}{m}\frac{(m-1)_3}{(n-1)_3}\frac{8p(G)}{n}$. Meanwhile, we can expand the square of the mean and obtain coefficients with the exact same summand terms $(\mathrm{E}[d(G^{|S})|G])^2 = \frac{1}{n^2}\left(\frac{m-1}{n-1}\right)^2\left(\sum_{i,j} A_{ij} + \sum A_{ij}A_{kl}\right)$. We have three sets of coefficients. Working with the coefficients on $\sum_{i,j} A_{ij}$, we have that

$$\begin{aligned}\frac{1}{m}\frac{(m-1)}{(n-1)}\frac{2|E|}{n} - \frac{1}{n}\left(\frac{m-1}{n-1}\right)^2\frac{2|E|}{n} &= \frac{1}{n-1}\left(\frac{m-1}{m} - \frac{(m-1)^2}{n(n-1)}\right)d(G) \\ &= \left(n^{-1}(1-\psi^2) + \Theta(n^{-2})\right)d(G).\end{aligned}$$

The second term is given by

$$\begin{aligned}\left(\frac{1}{m}\frac{(m-1)_2}{(n-1)_2} - \frac{1}{n}\left(\frac{m-1}{n-1}\right)^2\right)\frac{8t(G)}{n} &= \frac{1}{n-1}\left(\frac{1}{m}\frac{(m-1)(m-2)}{(n-2)} - \frac{(m-1)^2}{n(n-1)}\right)8t(G) \\ &= \left(n^{-1}\psi(1-\psi) + \Theta(n^{-2})\right)8t(G).\end{aligned}$$

Finally, we compute the last term $\left(\frac{1}{m}\frac{(m-1)_3}{(n-1)_3} - \frac{1}{n}\left(\frac{m-1}{n-1}\right)^2\right)\frac{8p(G)}{n} = \Theta(n^{-2})p(G)$, which completes the proof. ■

LEMMA D.2. *Under uniform random sampling of m out of n nodes, with $m/n \to \psi$ and as $n \to \infty$,*

$$\mathrm{E}[d(G^S)|G] = \psi(2-\psi+\Theta(n^{-1}))d(G) \text{ and } \mathrm{var}\left(d(G^S)|G\right) = \Theta(n^{-1}d(G))+\Theta(n^{-1}t(G))+\Theta(n^{-2}p(G)).$$

Proof. Same as previous lemma, noting each pair ij appears $|\{\sigma \in \mathcal{S} : i \vee j \in \sigma\}| = 2\binom{n-1}{m-1} - \binom{n-2}{m-2}$ times in $\mathcal{S}$. ■

LEMMA D.3. *Under uniform random sampling of m out of n nodes, with $m/n \to \psi$ and as $n \to \infty$,*

$$\frac{\rho(G^{|S})}{\tau(G^{|S})} = \frac{\rho(G)}{\tau(G)} + o(1).$$

Proof. Since the sampling probability is $\psi^3+o(1)$ for both the numerator and the denominator, there is no bias asymptotically in the estimate of clustering. ■

LEMMA D.4. *Under uniform random sampling of m out of n nodes, with $m/n \to \psi$ and as $n \to \infty$,*

$$\frac{\rho(G^S)}{\tau(G^S)} = \frac{\psi(3-2\psi)}{1+\psi(1-\psi)} \cdot \frac{\rho(G)}{\tau(G)} + o(1).$$

Proof. The expected number of triangles is given by $\mathrm{E}[\rho(G^S)|G] = (3\psi^2(1-\psi)+\psi^3+o(1))\rho(G)$ since with probability $3\psi^2(1-\psi)$ a transitive triangle can be counted with two nodes being sampled and ψ^3 with all three being sampled.[54] Meanwhile, $\mathrm{E}[\tau(G^S)|G] = (\psi(1-\psi)^2 + 3\psi^2(1-\psi)+\psi^3+o(1))\tau(G)$ where the extra term comes from the fact that the exact center of the triple in question can be selected. In turn,

$$\frac{\rho(G^S)}{\tau(G^S)} = \frac{3\psi^2(1-\psi)+\psi^3+o(1)}{\psi(1-\psi)^2+3\psi^2(1-\psi)+\psi^3+o(1)} \cdot \frac{\rho(G)}{\tau(G)} = \frac{\psi(3-2\psi)}{1+\psi(1-\psi)} \cdot \frac{\rho(G)}{\tau(G)} + o(1),$$

which proves the claim. ■

LEMMA D.5. *Under uniform random sampling of m out of n nodes, with $m/n \to \psi$ and as $n \to \infty$,*

$$\mathrm{E}\left[s\left(G^{|S}\right)|G\right] = \left\{1 - \sum_x (1-\psi)^x \, \mathrm{P}(x|G)\right\} s(G) + o(1)$$

and

$$\mathrm{E}\left[s\left(G^{S}\right)|G\right] = \frac{\psi^2 + 2\psi(1-\psi)\{1-\sum_x (1-\psi)^x \, \mathrm{P}(x|G)\}}{\psi^2+2(1-\psi)\psi} s(G) + o(1),$$

where $\mathrm{P}(x|G)$ *is the share of links ij that are supported by x in the underlying graph G.*

Proof. Start with the star subgraph. Observe there are three types of linked pairs: $i, j \in S$, $i \in S$ and $j \notin S$ (or vice vesa), and $i, j \notin S$. Obviously if neither are sampled, then observing ij is impossible. So we need only to check the rate of correctly observing a supported link

[54] We simply count, e.g., $\left\{3\binom{n-3}{m-2} + \binom{n-3}{m-3}\right\}\binom{n}{m}^{-1} = \frac{m(m-1)(m-2)}{n(n-1)(n-2)} = \psi^3 + o(1)$.

when either one or both are sampled. If both are sampled, clearly support is observed. So if only one is sampled, support is observed if and only if at least one of the supporting nodes is sampled. Let $x_{ij} = \sum_k A_{ik}A_{jk}$ be the number of supporting nodes of the pair ij. Then the probability of not observing support is $(1-\psi)^{x_{ij}}$. Since sampling is random, this implies that integrating over the distribution of the number of friends in common, the rate of support we observe among pairs where only one node is sampled is

$$1 - \sum_x (1-\psi)^x \mathrm{P}(x|G)$$

where $\mathrm{P}(x|G)$ is the share of links ij that are supported by x in the underlying graph G. Putting this together we have

$$\mathrm{E}\left[s\left(G^S\right)|G\right] = \frac{\psi^2 + 2\psi(1-\psi)\left\{1-\sum_x (1-\psi)^x \mathrm{P}(x|G)\right\}}{\psi^2 + 2(1-\psi)\psi} s(G) + o(1),$$

which proves the result.

Next turn to the induced subgraph and note here we correctly score support if and only if at least one of the nodes in common are also sampled. Therefore

$$\mathrm{E}\left[s\left(G^{|S}\right)|G\right] = \left\{1 - \sum_x (1-\psi)^x \mathrm{P}(x|G)\right\} s(G) + o(1)$$

which also proves the result. ∎

Lemma D.6. *Put $k(\psi) = \psi + \psi^2 - \psi^3$. For any sequence of random graphs $(G_n)_{n\in\mathbb{N}}$ satisfying, as $n\to\infty$, $d(G)/a_{1n} \xrightarrow{\mathrm{P}} c_1$, $d_2(G)/a_{2n} \xrightarrow{\mathrm{P}} c_2$, $a_{1n}, a_{2n} \in o(n)$, and $c_1, c_2 > 0$,*

(1) $\left|d(G^{|S}) - \psi d(G)\right| = o_{\mathrm{P}}(1)$, $\left|d_2(G^{|S}) - \psi^2 d_2(G)\right| = o_{\mathrm{P}}(1)$,
(2) $\left|d(G^S) - (1-(1-\psi)^2)d(G)\right| = o_{\mathrm{P}}(1)$, $\left|d_2(G^S) - k(\psi)d_2(G)\right| = o_{\mathrm{P}}(1)$.

This observation is general in the sense that it only requires that degree and the number of second neighbors to grow sufficiently slowly, which is reasonable for realistic applications.

Proof of Lemma D.6. We show $\left|d(G^S) - (1-(1-\psi)^2)d(G)\right| = o_{\mathrm{P}}(1)$. The arguments for $G^{|S}$ and $d_2(\cdot)$ are analogous. We have already shown that $\mathrm{E}\left[d(G^S)|G\right] = (1-(1-\psi)^2 + \Theta(n^{-1}))d(G)$ in Lemma D.2. Let χ_{ij} be an indicator of $i \vee j \in S$. Condition on the sequence of events $\mathcal{E}_n := \{d(G) \in (c_1 a_{1n} \pm \epsilon_1 a_{1n}) \cap d_2(G) \in (c_2 a_{2n} \pm \epsilon_2 a_{2n})\}$. By assumption of growth rates a_{1n} and a_{2n}, we these events happen with probability approaching one,

$$\mathrm{P}\left(\left|\frac{d(G)}{a_{1n}} - c_1\right| < \epsilon_1\right) \to 1 \text{ and } \mathrm{P}\left(\left|\frac{d_2(G)}{a_{2n}} - c_2\right| < \epsilon_2\right) \to 1.$$

Notice that the probability distribution and $d(G)$ are both implicitly indexed by n. Then using Chebyshev's inequality and letting $\varphi = (1-(1-\psi^2) + \Theta(n^{-1}))$,

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_i\sum_j A_{ij}\chi_{ij} - \varphi d(G)\right| > \epsilon \middle| \mathcal{E}_n\right) = \mathrm{P}\left(\left|\frac{1}{n}\sum_i\sum_{j>i} A_{ij}(\chi_{ij}-\varphi)\right| > \epsilon/2 \middle| \mathcal{E}_n\right) \le \frac{4}{n^2\epsilon^2}\mathrm{var}\left(\sum_i\sum_{j>i} Z_{ij}|\mathcal{E}_n\right)$$

where $Z_{ij} = A_{ij}(\chi_{ij} - \varphi)$. If we condition on a graph $G \in \mathcal{E}_n$,

$$\text{var}\left(\sum_i \sum_{j>i} Z_{ij}|G\right) = |E\left(G\right)| \,\text{var}\left(Z_{ij}|G\right) + \sum_i \sum_{j>i} \sum_{k \neq i,j} \text{cov}(Z_{ij}, Z_{jk}|G),$$

since $\text{var}(Z_{ij}|G) = \varphi(1-\varphi)$ and $\text{cov}(\chi_{ij}, \chi_{i'j'}|G) = \psi^3(1-\psi)$, with the covariance terms only entering when ij and $i'j'$ share a vertex. Recall that $d(G) = 2\left|E(G)\right|/n$, we have

$$\text{P}\left(\left|\frac{1}{n}\sum_i \sum_j A_{ij}\chi_{ij} - \varphi d(G)\right| > \epsilon \middle| \mathcal{E}_n, G\right) \lesssim \frac{1}{n}\left\{d\left(G\right)\varphi(1-\varphi) + d_2\left(G\right)\psi^3\left(1-\psi\right)\right\}$$

for every $G \in \mathcal{E}_n$ where the constant is uniform. Notice that $d(G) \in (c_1 a_{1n} \pm \epsilon_1 a_{1n})$ implies $d(G)/n \to 0$ because by assumption $a_{1n}/n \to 0$. The same is true for $a_{2n}/n \to 0$. This proves the result.

Next we look at $\left|d_2(G^S) - k(\psi)d_2(G)\right| = o_{\text{P}}(1)$.

$$\text{E}[n^{-1}\textstyle\sum_i \sum_{j\neq i} \sum_{k\neq i,j} A_{ij}A_{jk}\chi_{ij}\chi_{jk}|G] \quad = \quad n^{-1}\textstyle\sum_i \sum_{j\neq i} \sum_{k\neq i,j} A_{ij}A_{jk}\text{E}\left[\chi_{ij}\chi_{jk}|g\right] = k(\psi)d_2(G).$$

This follows from $\text{E}[\chi_{ij}\chi_{jk}|G] = 1 - (1-\psi)^2\left[2 - (1-\psi)\right]$. A similar argument to the above completes the proof. ■

Lemma D.7. $\text{E}\left[\left\|\bar{T}\right\|_F^2\right] = \|T\|_F^2 + \sum_i \xi_1(d_i, \psi)$, *where*

$$\xi_1(d_i, \psi) := \frac{1}{1-(1-\psi)^{d_i}} \sum_{r=1}^{d_i} \frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i - r + 1} - (1-\psi)/d_i.$$

Proof. Observe that $\left\|\bar{T}\right\|_F^2 = \sum_{i=1}^n \sum_{k=1}^n \bar{T}_{ik}^2 = \sum_{i:\bar{d}_i>0} \bar{d}_i^{-1}$ since $\bar{T}_{ik}^2 = \bar{d}_i^{-2}$ is greater than zero exactly $\bar{d}_i$ times. Note that $\text{E}\left[\bar{d}_i^{-1}|i \notin S, \bar{d}_i > 0\right]$ is the conditional expectation of the first negative moment of a binomial $\text{Bin}(d_i, \psi)$, namely[55]

$$\text{E}\left[\bar{d}_i^{-1}|i \notin S, \bar{d}_i > 0\right] = \frac{1}{1-(1-\psi)^{d_i}} \sum_{r=1}^{d_i} \frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i - r}.$$

The result follows from the fact that

$$\begin{aligned}\psi\frac{1}{d_i} + (1-\psi)\frac{1}{1-(1-\psi)^{d_i}} \sum_{r=1}^{d_i} \frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i - r} &= \frac{1}{d_i} + (1-\psi)\,\frac{1}{1-(1-\psi)^{d_i}} \sum_{r=1}^{d_i} \frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i - r} \\ &\quad - \frac{(1-\psi)}{d_i} \\ &= \frac{1}{d_i} + \xi_{1i}.\end{aligned}$$

Summing over the i nodes yields the result. ■

[55]See, e.g., Stephan (1945).

Lemma D.8. *$\mathrm{E}\left[\left\langle T,\bar{T}\right\rangle_F\right]=\|T\|_F^2+\sum_i\xi_2(d_i,\psi)$, where $\xi_2(d_i,\psi):=-d_i^{-1}(1-\psi)^{d_i+1}$.*

Proof. We can wite $\left\langle T,\bar{T}\right\rangle_F=\sum_{i=1}^n\sum_{k=1}^n T_{ik}\bar{T}_{ik}=\sum_{i:\bar{d}_i>0}\bar{d}_i^{-1}d_i^{-1}\bar{d}_i=\sum_{i:\bar{d}_i>0}d_i^{-1}$. As $\mathrm{P}\left(\bar{d}_i>0\right)=1-(1-\psi)^{d_i+1}$, where none of the d_i neighbors nor i is sampled, $\mathrm{E}\left[\left\langle T,\bar{T}\right\rangle_F\right]=\sum_i\frac{1-(1-\psi)^{d_i+1}}{d_i}=\|T\|_F^2+\sum_i\xi_2(d_i,\psi)$. ■

Lemma D.9. *$\mathrm{E}\left[Tr(\bar{T}'\iota\iota'\bar{T})\right]=\|T\|_F^2+\sum_i\xi_1(d_i,\psi)+\xi_3(\vec{d},\psi)$, where $\xi_3(\vec{d},\psi)$ is defined in the proof.*

Proof. Notice that

$$\begin{aligned}
\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T}) &= \sum_i\left|\left[\iota'\bar{T}\right]_i\right|^2\\
&= \sum_i\left(\sum_k\bar{T}_{ki}^2+2\sum_{l>k}\sum_k\bar{T}_{ki}\bar{T}_{li}\right)\\
&= \sum_{k:\bar{d}_k>0}\bar{d}_k^{-1}+2\sum_{l>k}\sum_k\sum_i\bar{T}_{ki}\bar{T}_{li}\\
&= \left\|\bar{T}\right\|_F^2+2\sum_{l>k:\bar{d}_l>0}\sum_{k:\bar{d}_k>0}\frac{\left|\bar{N}_k\cap\bar{N}_l\right|}{\bar{d}_k\bar{d}_l}.
\end{aligned}$$

With probability ψ^2 both $k\in S$ and $l\in S$, so $\bar{d}_k^{-1}\bar{d}_l^{-1}\left|\bar{N}_k\cap\bar{N}_l\right|=d_k^{-1}d_l^{-1}\left|N_k\cap N_l\right|=:c(k,l)$. With probability $\psi(1-\psi)$ we have $\zeta_4(k,l)$ and with the same probability we have $\zeta_4(l,k)$, where ζ_4 is defined in Lemma D.10 below. Finally, with probability $(1-\psi)^2$

$$\zeta_3(k,l):=\sum_{r=1}^{N_l-|N_k\cap N_l|}\sum_{t=1}^{N_k-|N_k\cap N_l|}\sum_{s=1}^{|N_k\cap N_l|}\binom{N_l-|N_k\cap N_l|}{r}\binom{N_k-|N_k\cap N_l|}{t}\binom{|N_k\cap N_l|}{s}\frac{\psi^{s+t+r}(1-\psi)^{|N_k\cup N_l|-s-t-r}}{(t+s)(r+s)}.$$

Then $\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T})\right]=\|T\|_F^2+\sum_i\xi_1(d_i,\psi)+\xi_3(\vec{d},\psi)$ where

$$\xi_3(\vec{d},\psi):=2\sum_{l>k}\sum_k\left\{\psi^2c(k,l)+\psi(1-\psi)\zeta_4(l,k)+\psi(1-\psi)\zeta_4(k,l)+(1-\psi)^2\zeta_3(k,l)\right\}$$

which completes the proof. ■

Lemma D.10. *$\mathrm{E}\left[Tr(\bar{T}'\iota\iota'T)\right]=\|T\|_F^2+\sum_i\xi_2(d_i,\psi)+\xi_4(\vec{d},\psi)$, where $\xi_4(\vec{d},\psi)$ is defined below.*

Proof. We have

$$\mathrm{Tr}(\bar{T}'\iota\iota'T)=\sum_i[\bar{T}'\iota]_i\left[\iota'T\right]_i=\sum_i\left(\sum_k\bar{T}_{ki}T_{ki}+2\sum_{l>k}\sum_k\bar{T}_{ki}T_{li}\right)=\left\langle T,\bar{T}\right\rangle_F+2\sum_{l>k}\sum_{k:\bar{d}_k>0}\sum_i\bar{T}_{ki}T_{li}.$$

The first term has already been controlled in Lemma D.8. To compute the second term, observe that with probability ψ, $k\in S$ and therefore and in this case $\sum_i\bar{T}_{ki}T_{li}=c(k,l)$.

With probability $1-\psi$, $k \notin S$, and as such the conditional expectation is given by

$$\zeta_4(k,l) := \sum_{t=1}^{N_k - |N_k \cap N_l|} \sum_{s=1}^{|N_k \cap N_l|} \binom{|N_k \cap N_l|}{s} \binom{N_k - |N_k \cap N_l|}{t} \frac{\psi^{s+t}(1-\psi)^{N_k - s - t}}{(s+t)d_l}.$$

So $\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota' T)\right] = \|T\|_F^2 + \sum_i \xi_2(d_i,\psi) + \xi_4(\vec{d},\psi)$ where $\xi_4(\vec{d},\psi) = 2\sum_{l>k}\sum_k \{\psi c(k,l) + (1-\psi)\zeta_4(k,l)\}$ which completes the proof. ■

Appendix E. Supplementary Proofs for Section 4

E.1. **Useful Results.** In what follows, for $p \times q$ matrix D, let $\|\cdot\| := \|\cdot\|_2$ be the matrix norm induced by vector norm $\|\cdot\|$, with $\|D\| := \max_{x \in S^{q-1}} \|Dx\|$.

Lemma E.1. *Assume for all $\theta_r \in \Theta_r$, $\zeta_{ir}(\theta_r)$ are $q \times kq$ matrix (or vector) valued functions satisfying $\sup_{i,r} \|\zeta_{ir}(\theta_r)\| \le B_{ir}$ with $\limsup_{R\to\infty} \mathbb{E}_{nR}[\mathrm{E}B_{ir}] < \infty$ and Assumption A.3.3 hold. Then*

$$(nR)^{1/2} \left\| \mathbb{E}_{n,R}\left[\zeta_{ir}(\bar{\theta}_r)(I_q \otimes (\widehat{\theta}_r - \theta_{0r}))\right] \right\| = o_{\mathrm{P}}(1)$$

for $\bar{\theta}_r$ on the line between $\widehat{\theta}_r$ and θ_{0r}.[56]

Proof of Lemma E.1. This follows from

$$(nR)^{1/2}\mathbb{E}_{n,R}\left[\left\|\zeta_{ir}(\bar{\theta}_r)\right\| \left\|I_q \otimes (\widehat{\theta}_r - \theta_{0r})\right\|\right] \le (\mathbb{E}_{nR}[B_{ir}]) \cdot \sqrt{nR} \sup_{r \le R} \left\|\widehat{\theta}_r - \theta_{0r}\right\|$$
$$\le \mathbb{E}_{nR}[B_{ir}] \cdot O_{\mathrm{P}}\left(a_n^{-1} \cdot \sqrt{nR^{1+2/b}}\right) = o_{\mathrm{P}}(1)$$

as the mean of the envelopes converges and the rates obey the assumed relationship. ■

Lemma E.2. *Conditions 1 and 2 of Lemma 4.1 imply* $\sup_r \left\|\widehat{\theta}_r - \theta_{0r}\right\| = o_{\mathrm{P}}(1)$.

Proof of Lemma E.2. Arguing along the lines of Theorem 5 of Supplementary Appendix I of Hahn and Newey (2004), among others, pick $\eta > 0$, define $\epsilon := \inf_{r \le R}\left[Q_{(r)}(\theta_{0r}) - \sup_{\theta_r : \|\theta_r - \theta_{0r}\| > \eta} Q_{(r)}(\theta_r)\right]$, and condition on the event $\left\{\sup_{r\le R}\sup_\theta \left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| < \frac{\epsilon}{3}\right\}$ which has probability $1 - o(a_n^{-v})$ by Condition 2. If we look for $\widehat{\theta}_r$ outside an η-radius ball of the true parameter, we have $\sup_{\theta_r:\|\theta_r-\theta_{0r}\|>\eta} \widehat{Q}_{(r)}(\theta_r) < \sup_{\theta_r:\|\theta_r-\theta_{0r}\|>\eta} Q_{(r)}(\theta_r) + \frac{\epsilon}{3} < Q_{(r)}(\theta_{0r}) - \frac{2\epsilon}{3} < \widehat{Q}_{(r)}(\theta_{0r}) - \frac{\epsilon}{3}$, contradicting $\widehat{Q}_{(r)}(\widehat{\theta}_r) \ge \widehat{Q}_{(r)}(\theta_{0r})$, implying $\left\|\widehat{\theta}_r - \theta_{0r}\right\| < \eta$ for all networks. ■

The next lemma is useful throughout and we make explicit the dependence on R in $\mathrm{P}^R(\cdot)$ and E^{R} here.

[56] Each component of $\bar{\theta}_r$ may be at a different intermediate point.

Lemma E.3 (Extended Vitali Convergence). *Let $\{Z_R : R \in \mathbb{N}\}$ be L^1_R-integrable functions on a sequence of measure spaces indexed by R. (1) $Z_R \xrightarrow{\mathrm{P}} 0$ and (2) Z_R is uniformly integrable, $\sup_{R\geq 1} \mathrm{E}^R\left[|Z_R| \cdot \mathbf{1}\{|Z_R| > c\}\right] \to 0$ as $c \to \infty$, imply $\mathrm{E}^R[Z_R] \to 0$.*

Proof. The argument is analogous to the proofs of Theorem 10.3.5 and 10.3.6 in Dudley (2002). Let $Z_R \geq 0$ w.l.o.g. First, observe (2) implies that for every $\epsilon > 0$ there exists $\delta > 0$ such that for each A_R with $\mathrm{P}^R(A_R) < \delta$, $\mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\}] < \epsilon$ for all R. To see this, by uniform integrability, given $\epsilon > 0$ and pick $\delta < \epsilon/(2K)$ where K is large enough such that $\mathrm{E}^R[Z_R \cdot \mathbf{1}\{Z_R > K\}] < \epsilon/2$. Then

$$\mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\}] \leq \mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\} \cdot \mathbf{1}\{Z_R \leq K\}] + \mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\} \cdot \mathbf{1}\{Z_R > K\}] < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

as $\mathrm{P}^R(A_R) < \delta$. Second, given $\epsilon > 0$, let $A_R := \{Z_R > \epsilon\}$. By (1), $\mathrm{P}^R(A_R) \to 0$ as $R \to \infty$. As such, for R large enough $\mathrm{P}^R(A_R) < \epsilon/(2K)$. Therefore $\mathrm{E}^R[Z_R] = \mathrm{E}^R[Z_R \cdot \mathbf{1}\{Z_R \leq \epsilon\}] + \mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\}] \leq 2\epsilon$. ∎

Lemma E.4. *Let $Y_{ir,R}$ be a triangular array of mean-zero random variables, $Z_{ir,R} := \mathrm{E}[Y_{ir,R}|X_r]$, and define*

$$\lambda_r := \sqrt{n}\mathbb{E}_n[Y_{ir}] \text{ and } \zeta_r := \sqrt{n}\mathbb{E}_n[Z_{ir}].$$

If (1) $\sup_R \sup_r \mathrm{var}(\lambda_r) < C_1$ and (2) $\inf_R \inf_r \mathrm{var}(\zeta_r) > C_0$, then $\sqrt{nR}\mathbb{E}_{n,R}[Z_{i,r}] \rightsquigarrow \mathcal{N}(0, V_Z)$.

Proof. Identical to the proof of step 1 in "Proof that Lemma 11 of Chernozhukov et al. (2009) applies". ∎

E.2. Supplementary Proofs for Example 1.

Lemma E.5. *Under Assumption A.7.3, $\sup_r \left|\frac{1}{\sqrt{|\Xi|}} \sum_{s\in\Xi} v(X_{rs};\theta_{0r})\right| = O_{\mathrm{P}}(R^{1/b})$.*

Proof. Let $T := |\Xi|$ and define $y_{rT} := \sum v(X_{rs};\theta_{0r})/\sqrt{T}$. We assume y_{rT} has b^{th} moments, $\sup_{r\leq R} \mathrm{E}\|y_{rT}\|^b < \infty$, which will be shown below. That $\left(\mathrm{E}\|\sup_r y_{rT}\|^b\right)^{1/b} \leq R^{1/b}\sup_r \left(\mathrm{E}\|y_{rT}\|^b\right)^{1/b}$ implies $\sup_r y_{rT} = O_{\mathrm{P}}(R^{1/b})$. Notice $\mathrm{E}\left[\|y_{rT}\|^b\right] = \mathrm{E}\left[\left\|\sum_s v(X_{rs};\theta_{0r})\right\|^b\right] T^{-b/2}$. We need only control $\mathrm{E}\left[\left\|\sum_s v(X_{rs},\theta_{0r})\right\|^b\right] \leq \mathrm{E}\left[\left(\sum_s B(X_{rs})\right)^b\right]$. Observe $\mathrm{E}\left[\left(\sum_{s\in\Xi} B(X_{rs})\right)^b\right] = \mathrm{E}\left[\sum B(X_{rs})^b\right] + \mathrm{E}\left[\sum\prod_j B(X_{rs_j})^{\gamma_j}\right]$ where $\sum\gamma_j = b$. If 2^{b-1} moments exist for the envelopes, then this can be majorized into terms of $\prod \mathrm{E}[B(X_{rs_j})^{\delta_j}]$ where $\delta_j \leq 2^{b-1}$ by repeated application of Holder inequalities. ∎

Lemma E.6. *Under Assumptions A.7 and A.6, $\mathrm{P}\left(\sup_r \sup_{\theta\in\Theta}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right) = o\left(|\Xi|^{-1}\right)$.*

Proof. The argument is along the lines of Lemma 2 of Hall and Horowitz (1996) and Lemma 3 of Supplementary Appendix I of Hahn and Newey (2004). We use Lemma E.7. First we use a union bound over the R graphs and focus on $\sum_{r=1}^{R} \mathrm{P}\left(\sup_{\theta\in\Theta}\left|\widehat{Q}_{(r)}(\theta)(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right)$. Next, considering a given graph, we choose $\epsilon > 0$ such that $2\epsilon \cdot \sup_r \mathrm{E}[B(X_{rs})] < \frac{\eta}{3}$. Divide Θ into subsets $\Theta_1, ..., \Theta_{M(\epsilon)}$ such that $\|\theta - \theta'\| \leq \epsilon$ when θ and θ' are in the same subset. A second union bound gives us $\sum_{j=1}^{M(\epsilon)} \mathrm{P}\left(\sup_{\theta\in\Theta_j}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right)$.

Let θ_j denote a point in Θ_j. Noticing that,

$$\begin{aligned}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| &\leq \left|\widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta_j)\right| + \left|\widehat{Q}_{(r)}(\theta) - \widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta) + Q_{(r)}(\theta_j)\right| \\ &\leq \left|\widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta_j)\right| + \frac{\epsilon}{|\Xi|}\left|\sum B(X_{rs}) - \mathrm{E}[B(X_{rs})]\right| + 2\epsilon \mathrm{E}[B(X_{rs})],\end{aligned}$$

and

$$\begin{aligned}\mathrm{P}\left(\sup_{\theta\in\Theta_j}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right) &\leq \mathrm{P}\left(\left|\widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta_j)\right| \geq \frac{\eta}{3}\right) \\ &\quad + \mathrm{P}\left(|\Xi|^{-1}\left|\sum_{s\in\Xi}\left(B(X_{sr}) - \mathrm{E}B(X_{sr})\right)\right| \geq \frac{\eta}{3\epsilon}\right) = o(|\Xi|^{-k})\end{aligned}$$

by Lemma E.7, the result holds as $R = o(|\Xi|^k)$. ■

Lemma E.7. *For each r, suppose $\{X_{sr} : s \in \Xi\}$ be covariates satisfying Assumption A.6. Let $R = O\left(|\Xi|^h\right)$, and let h, k, p, γ, with k, p, γ defined below, satisfy $h + 1 \leq k < p/2 - \gamma pd$. Then $\forall\eta > 0$,*

$$\max_r \mathrm{P}\left(\left||\Xi|^{-1}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) = o\left(|\Xi|^{-k}\right) \text{ and } \mathrm{P}\left(\max_r\left||\Xi|^{-1}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) = o\left(|\Xi|^{-1}\right).$$

Proof. The argument follows Lemma 1 of Hall and Horowitz (1996) and Lemma 2 of Supplementary Appendix I of Hahn and Newey (2004).

Step 1: By Chebyshev's inequality

$$\mathrm{P}\left(\left||\Xi|^{-1}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) \leq \frac{C \cdot \mathrm{E}\|X_{1r}\|^{p+\delta}\left(|\Xi|^{p/2} m^{dp} + |\Xi|^p \alpha^r_{2,\infty}(m)^{\frac{\delta}{p+\delta}}\right)}{\eta^p |\Xi|^p},$$

for $1 \leq m \leq C(p)|\Xi|$ where the second inequality follows from Lemma E.8, which we can write under Assumption A.6.[57] We can bound the right hand side by $\eta^{-p} C \cdot \mathrm{E}\|X_{1r}\|^{p+\delta} m^{dp} |\Xi|^{-p/2} + \eta^{-p} C \cdot \mathrm{E}\|X_{1r}\|^{p+\delta} \alpha^r_{2,\infty}(m)^{\frac{\delta}{p+\delta}}$ and for $m = |\Xi|^\gamma$ for some γ with $0 < \gamma \leq 1$, using the bound

[57] We write the proof for $\sup_r \alpha^r_{\infty,2}(m) \leq Ca^m$, though the extension to $\sup_r \alpha^r_{\infty,2}(m) = o\left(m^{-d}\right)$ is straightforward and will merely result in more stringent requirements on r, k, γ, d. For Assumption A.6(ii) we have

$$\max_r |\Lambda|^k \mathrm{P}\left(\left|\frac{1}{|\Xi|}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) = O\left(|\Xi|^{d\gamma p + k - p/2} + |\Xi|^{k - \gamma d - \gamma\epsilon}\right)$$

for $\alpha^r_{1,\infty}(m) = O\left(m^{-d-\epsilon}\right) = o\left(m^{-d}\right)$. Then the requirement is $k < (p/2 + \gamma pd) \vee (d + \epsilon)$.

$\sup_r \alpha^r_{\infty,1}(m) \le Ca^m$ on the mixing coefficient,

$$\max_r |\Xi|^k \, \mathrm{P}\left(\left||\Xi|^{-1}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) \le \eta^{-p} C \cdot \max_r \mathrm{E}\left\|X_{1r}\right\|^{p+\delta}\left(|\Xi|^{\gamma pd+k-p/2} + |\Xi|^k \, a^{\frac{\delta}{p+\delta}|\Xi|^{\gamma}}\right).$$

We have shown that for Assumption A.6, $\max_r |\Xi|^k \, \mathrm{P}\left(\left|\frac{1}{|\Xi|}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) = O\left(|\Xi|^{\gamma pd+k-p/2}\right) = o(1)$ if $\gamma pd + k < p/2$ which is the first result.

Step 2: By a union bound $|\Xi| \, \mathrm{P}\left(\max_r \left|\frac{1}{|\Xi|}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) \le |\Xi| \, R \cdot \max_r \mathrm{P}\left(\left|\frac{1}{|\Xi|}\sum_{s\in\Xi} X_{sr}\right| > \eta\right)$. If $R = O\left(|\Xi|^h\right)$, it follows that $|\Xi| \, \mathrm{P}\left(\max_r \left|\frac{1}{|\Xi|}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) \le O\left(|\Xi|^{h+1}\right) o\left(|\Xi|^{-k}\right)$. Since $h+1 \le k$, $O\left(|\Xi|^{h+1}\right) o\left(|\Xi|^{-k}\right) = o(1)$, $\mathrm{P}\left(\max_r \left|\frac{1}{|\Xi|}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) = o\left(|\Xi|^{-1}\right)$ which proves the result. ■

Lemma E.8. *Let $\left\{X_i : t_i \in \Lambda \subset \mathbb{Z}^d\right\}$ be a mean zero stationary random field satisfying Assumption A.6 and $\{X_{ij} : ij \in \Xi\}$ covariates. Then for any positive integer r and for $1 \le m < C(k) \cdot |\Xi|$, we have* $\mathrm{E}\left[\left(\sum_{ij\in\Xi} X_{ij}\right)^k\right] \le C(r)\,\mathrm{E}\left\|X_1\right\|^{k+\delta}\left(|\Xi|^{k/2} m^{kd} + |\Xi|^k \alpha_{2,\infty}(m)^{\frac{\delta}{k+\delta}}\right)$.

Proof. The proof builds on Lahiri (1992), with two differences: the first is an extension to mixing random fields,[58] and the second is that we are interested over moments of random variables on Ξ as opposed to Λ. The Lahiri (1992) style of argument proceeds in four parts; we include the entire argument for completeness though the key differences are in the last two steps. First, we can control the first $k/2$ terms via a standard result. This will enable us to bound this part of the sum by a $|\Xi|^{k/2}$ rate. Second, for the remaining terms, we will divide the space into a set of all pieces with a well separated point τ on the lattice whose random variable X_τ has power 1 and is at least of distance m from any other point in the collection, and into its complement. Third, we will control this set using the mixing coefficient and fourth, by a counting argument we create an upper bound on the number of points in the complement. It is useful to note that for $ij \in \Xi$, for $1 \in \Lambda$, $\mathrm{E}\left\|X_{ij}\right\|^k \lesssim \mathrm{E}\left\|X_1\right\|^k$ by the assumption and stationarity. Moreover, we defined the pseudo-metric $d_\Xi(ij, kl) := d_\Lambda(i,k) \wedge d_\Lambda(i,l) \wedge d_\Lambda(j,k) \wedge d_\Lambda(j,l)$ where $d_\Lambda(x,y) := \|x-y\|_1$.

Step 1: For $k = 2h$, we can expand the term into a polynomial,

$$\left(\sum_{s\in\Xi} X_s\right)^k = \sum_{j=1}^k \sum_{\alpha_1,\ldots,\alpha_j} c(\alpha_1,\ldots,\alpha_j) \sum_{s_1,\ldots,s_j} \prod_{t=1}^j X_{s_t}^{\alpha_t}$$

where $t = 1,\ldots,j$ is an arbitrary index of a j-tuple $(s_1,\ldots,s_j) \subset \Xi$ and $c(\cdot)$ are coefficients. We can control the first $k/2$-tuples by a standard argument, e.g., Bhattacharya and Rao (1986), making use of $\mathrm{E}\left\|X_{ij}\right\|^k \lesssim \mathrm{E}\left\|X_1\right\|^k$,

$$\left|\sum_{j=1}^{k/2} \sum_{\alpha_1,\ldots,\alpha_j} c(\alpha_1,\ldots,\alpha_j) \sum_{s_1,\ldots,s_j} \mathrm{E}\left[\prod_{t=1}^j X_{s_t}^{\alpha_t}\right]\right| \le C(k)\,|\Xi|^{k/2}\,\mathrm{E}\left\|X_1\right\|^k.$$

[58] Hahn and Kuersteiner (2004) extend the Lahiri argument to the time series setting.

In what follows it suffices to show for fixed $j > k/2$ and $(\alpha_1, ..., \alpha_j)$,

$$\left|\sum_{s_1,...,s_j} \mathrm{E}\left[\prod_{t=1}^{j} X_{s_t}^{\alpha_t}\right]\right| \leq C(k)\,\mathrm{E}\left\|X_1\right\|^{k+\delta}\left(|\Xi|^{k/2} m^{kd} + |\Xi|^{k}\alpha_{2,\infty}(m)^{\frac{\delta}{k+\delta}}\right).$$

Step 2: Next, we create a set that counts the sites s_τ where X_{s_τ} has power $\alpha_\tau = 1$ and s_τ is sufficiently far from the other s_t in the j-tuple. Let $u := j - k/2$. We put $A := \{t : \alpha_t = 1\}$ as the set of all points that have coefficient 1. Then let $\beta_0 = |A|$. This is a set that counts the number of indices that show up exactly one time. We want to show that this set is non-empty. Note that $1 \leq u \leq k/2$. Also, since $k = \sum_{t=1}^{j}\alpha_t \geq \beta_0 + 2(j - \beta_0)$, this means $2u \leq \beta_0 \leq k$ and therefore $(j - k/2) \leq \beta_0 \leq k/2$.

Then we partition the set of all j-tuples into B_m and B_m^c, with In Lahiri's notation, $\sum_2 \equiv \sum_{s_1,...,s_j} =: \sum_3 + \sum_4$. To define the sets, put

$$B_m := \{(s_1, ..., s_j) : \inf_{l \neq t} d_\Xi(s_l, s_t) = d_\Lambda(i_{s_l}, k_{s_t}) \wedge d_\Lambda(j_{s_l}, k_{s_t}) \wedge d_\Lambda(i_{s_l}, l_{s_t}) \wedge d_\Lambda(i_{s_l}, l_{s_t}) > m \text{ for some } t \in A\}.$$

Now $\sum_3$ sums over the terms in B_m and $\sum_4$ over the terms in B_m^c.

Step 3: We want to control B_m. Fix $\tau \in A$. Then decompose $\prod_{t=1}^{j} X_{s_t}^{\alpha_t} = X_a X_b$, with $X_b = X_{s_\tau}$ and $X_a = \prod_{t \neq \tau} X_{s_t}^{\alpha_t}$. Then X_a is a random field with respect to $\sigma\left(X_{s_1}, ..., \widehat{X}_{s_\tau}, ..., X_{s_j}\right)$, where we use notation to indicate the omission of a term, $\left(a, \widehat{b}, c\right) := (a, c)$, and X_b with respect to $\sigma(X_{s_\tau})$. These are of size $j - 1$ and 1, respectively. By definition the distance between these two sets is at least m, so by applying Lin et al. (1996) and using the fact that $\mathrm{E}[X_{ij}X_{kl}] \leq \|X_{ij}\|_{L^p(\mathrm{P})} \cdot \|X_{kl}\|_{L^q(\mathrm{P})}\,\alpha_{2,2}^{1-q^{-1}-p^{-1}}(d_\Xi(ij, kl))$, we have

$$\left|\mathrm{E}\left[\prod_{t=1}^{j} X_{s_t}^{\alpha_t}\right]\right| \leq |\mathrm{E}[X_a X_b]| \leq C \cdot \|X_a\|_{L^p(\mathrm{P})} \cdot \|X_b\|_{L^q(\mathrm{P})}\,\alpha_{2(j-1),2}^{1-q^{-1}-p^{-1}}(m).$$

Taking $p = k + \delta$ and $\frac{1}{q} = \frac{k-1}{k+\delta}$, using stationarity and repeatedly applying the Holder inequality,

$$\left|\mathrm{E}\left[\prod_{t=1}^{j} X_{s_t}^{\alpha_t}\right]\right| \leq C \cdot \|X_a\|_{L^{k+\delta}(\mathrm{P})} \cdot \|X_b\|_{L^{(k-1)/(k+\delta)}(\mathrm{P})}\,\alpha_{2(j-1),2}^{\delta/(k+\delta)}(m) \leq C \cdot \mathrm{E}\left\|X_1\right\|^{k+\delta}\alpha_{2(j-1),2}^{\delta/(k+\delta)}(m).$$

Since $\alpha_{j,1}(m) \leq \alpha_{j+1,1}(m)$, $\left|\mathrm{E}\left[\sum_{B_m}\prod_{t=1}^{j} X_{s_t}^{\alpha_t}\right]\right| \leq C \cdot \mathrm{E}\left\|X_1\right\|^{k+\delta}|\Xi|^{k}\alpha_{2,\infty}(m)^{\frac{\delta}{k+\delta}}$.

Step 4: Finally, we control B_m^c. To get a (coarse) upper bound, first notice the maximum number of powers of 1 that can be placed is $2u$[59]. Construct a set $\Gamma \subset \{1, ..., j\}$, $|\Gamma| = 2u$ which will include all powers of one and perhaps some residual copies of terms with higher power. Then we will simply count

$$\bar{B}_m^c := \left\{(s_1, ..., s_j) : \inf_{l \neq t} d_\Xi(s_t, s_l) = d_\Lambda(i_{s_l}, k_{s_t}) \wedge d_\Lambda(j_{s_l}, k_{s_t}) \wedge d_\Lambda(i_{s_l}, l_{s_t}) \wedge d_\Lambda(i_{s_l}, l_{s_t}) \leq m\ \forall t \in \Gamma\right\}.$$

It will help us to define the *partners* of s_t as $\epsilon(t) \in \arg\inf_l d_\Xi(s_t, s_l)$. Then define $\mathcal{P}(\Gamma)$ as the set of partners of $t \in \Gamma$ and put $v = |\Gamma \cup \mathcal{P}(\Gamma)|$. This is the number of distinct elements of Ξ with power one or power greater than one whose copies form the residual members of

59 Let x be the maximal number. Then $k = x + 2(j - x)$. Solving this yields the result.

Γ and their partners. By definition $2u \le v \le 4u \wedge j$ since $v \ge |\Gamma|$ and $v \le 2|\Gamma|$ if each had a distinct partner.

We define an *m-unbroken set* as a collection of points in $\Gamma \cup \mathcal{P}(\Gamma)$ for which each member is within an m-distance under pseudo-metric d_Ξ. Put q as the number of m-unbroken sets. First, observe that there are less than $|\Xi|^q$ initial sites to place a seed for each of the unbroken sets.[60] Next, we have to place each of the $v-q$ terms. For a given ij there are less than $2(2m+1)^d$ elements of Ξ within an m-distance since $d_\Lambda(k,i) \wedge d_\Lambda(l,i) \wedge d_\Lambda(k,j) \wedge d_\Lambda(l,j) \le m$. An upper bound is $O\left(m^{(v-q)d}\right)$. Finally, we can arbitrarily place the remaining $j-v$ elements yielding $|\Xi|^{j-v}$. This gives us $\left|\bar{B}_m^c\right| \lesssim |\Xi|^{q+j-v} m^{(v-q)d}$. The same counting exercise as in Lahiri (1992) gives us $j+q-v \le h$ and $v-q \le k$ which yields $|\Xi|^{q+j-v} m^{(v-q)d} \le |\Xi|^h m^{kd}$. This concludes the proof. ∎

Appendix F. Supplementary Data Catalogue

This appendix contains a catologue of a number of papers using network data in the applied literature. We have included only one paper per data-set. Notice some of the publicly available data (such as Add Health or the Karnataka villages or the Townsend Thai Data) have been used numerous times by both the researchers who collected the data as well as other researchers, but that is not reflected here. We do not claim that this is representative of the entire body of work on networks, but we hope this communicates what typical applied micro network datasets may look like, especially in development.

Paper	**Implied Sampling Rate ψ**	**Note 1**	**Note 2**
Ali and Dwyer (2009)	0.2305	Top Coding	
Alatas et al. (2016)	0.683		
Bandiera and Rasul (2006)	0.0753		
Banerjee et al. (2013)	0.46		
Cai et al. (2015)	Not Enough Information	Top Coding	
Conley and Udry (2010)	0.25		
Fafchamps and Lund (2003)	0.3	Top Coding	Denominator Underestimated
Feigenberg et al. (2010)	0.003		
Goeree et al. (2010)	0.77		
Kinnan and Townsend (2012)	0.41		
Kremer and Miguel (2007)	Not Enough Information	Top Coding	
Leider et al. (2009a) (Dictator Game Sample)	0.71	Top Coding	
Leider et al. (2009a) (Helping Game Sample)	0.46		
Ligon and Schechter (2012)	0.396	Denominator Underestimated	
Marmaros and Sacerdote (2006)	0.429	Denominator Underestimated	
Ngatia (2011)	1		

Above, "Not Enough Information" means that we were unable to determine what the likely sampling rate was from the paper listed. "Top Coding" means that the authors limited the number of links a node could name (typically 5, though also 8 or 10). Notice that this presents an additional problem beyond what has been focused on in the present paper. In at least two

[60] We leave ourselves open to certain double-counting; the bound is coarse.

of these papers, top coding was at 5 links and the mean number of friends in the sample was 4.8 and 4.9, indicating that the typical individual had at least 5 links and therefore sampling in this manner generated nearly a 5-regular graph. “Denominator Underestimated” means that we were unable to determine the number of nodes per network in the sample. Typically in this case the authors report the size of their sample ($|S|$) and then, since they are using star sampling, the number of unique additional contacts. Using this we can compute what is likely an extreme and crude upper bound on the sampling rate: $|S|$ / ($|S|$ + additional contacts). Therefore in these cases we suspect that the true implied sampling rate is below the one we state in the table.

Appendix G. Overview of Estimation Algorithm and Standard Errors

We present a non-technical overview of the standard errors and estimation procedures used. A theoretical discussion of standard error estimation is beyond the scope of this paper. Here, we discuss standard errors and finite-sample simulation bias. A lengthier formal, technical discussion and simulations results are available from the authors upon request. Several approaches can be used for inference: heteroskedasticity-robust, clustered, block bootstrapped, first-stage bootstrapped, and importance sampled standard errors. Network-level inference can be performed using heteroskedasticity-robust standard errors under cross-network independence, even in the presence of sampling. Individual-level inference under sampling requires attention to within-network, cross-individual autocorrelation exacerbated by the reconstructed regressor's mismeasurement, what we call the reconstruction error. When an edge is missing in the computation of one individual's network statistic, that edge is also missing for all other individuals' network statistic computation. Clustered standard errors and their nonparametric block bootstrap analog both account for heteroskedasticity and within-graph autocorrelation. Additionally, a parametric bootstrap which simulates the autocorrelation from the reconstruction error can estimate the contribution of the sampling-induced autocorrelation to the reconstruction estimator's variance. The parametric structure of this bootstrap should allow these standard errors to be estimated on a smaller collection of networks than necessary for clustered standard errors.

While the graphical reconstruction estimator is formally a “two-step” estimator, superconsistency allows the researcher to ignore the first-stage estimation uncertainty. That said, first-stage uncertainty can be understood using a first-stage parametric bootstrap and importance sampling. The first-stage parametric bootstrap estimates the distribution of the first-stage parameters and encompasses the following effects: sample from the observed collection of networks with replacement, estimate each network's first-stage parameters using the observed data, simulate a new collection of observed graphs using the first-stage estimates, and estimate the first-stage parameters using the simulated graphs. Repeat this process to obtain the bootstrapped variance of the first stage estimators to determine whether super-consistency is justified. Importance sampling can reduce the computational burden of bootstrapping the first-stage estimates by using the first-stage model's likelihood to rebalance a collection of reconstructed network statistics and greatly reduces the cost of accounting for any first-stage variability.

With or without first-stage superconsistency, the reconstruction estimator still uses a finite number of simulations, leaving another error term which we call the simulation error. A regularization to estimate and subtract out the variance in the regressors due to this noise can be employed.

Appendix H. Discussion of Analytical Corrections with Regularization

H.1. **A Regularized Estimator.** We show a simple example where our analytical correction is consistent and does not require us to assume $\sigma_v^2 \to 0$. In section 3, for degree and graph clustering, we have shown that $\widetilde{\beta} = \Phi\widehat{\beta} \xrightarrow{\text{P}} \beta_0 \frac{\sigma^2_{\langle x\rangle}}{\sigma^2_{\langle x\rangle} + \Phi^{-2}\sigma_v^2}$ where Φ is a deterministic function of sampling rate ψ. We now show we can construct estimates $\widehat{\sigma}^2_{\langle x\rangle}$ and $\widehat{\sigma}^2_v$ and therefore develop a consistent analytical correction $\beta^{\star}$. We present a simple example to illustrate the argument. Our example is average degree and for simplicity assume the researcher samples every edge independently with probability ψ. It is easy to see $\mathrm{E}\left[d(\widetilde{G})|G\right] = \psi d(G)$, where $\bar{G}$ is the sampled graph. Putting $v = d(\widetilde{G}) - \psi d(G)$, we can analytically compute $\mathrm{var}\left(v|G\right) = \frac{2}{n}\psi\left(1-\psi\right)d(G)$.

Lemma H.1. $\mathrm{var}\left(v|G\right) = 2n^{-1}\psi\left(1-\psi\right)d(G)$.

Proof. Let $T = \binom{n}{2}$ and e index edges going from $e = 1, ..., \binom{n}{2} = T$. Notice average degree is $d(G) = \frac{2}{n}\sum_e A_e$. Then the $v = \frac{2}{n}\sum_e \chi_e A_e - \psi\frac{2}{n}\sum_e A_e$. Notice

$$\begin{aligned}
\mathrm{E}\left[\frac{2}{n}\sum\nolimits_e \chi_e A_e\right]^2 &= \mathrm{E}\left[4n^{-2}\left\{\sum\nolimits_e \chi_e A_e + 2\sum\nolimits_{e<e'}\chi_e\chi_{e'}A_eA_{e'}\right\}\right] = 4n^{-2}\left\{\psi\sum\nolimits_e A_e + \psi^2 2\sum\nolimits_{e<e'}A_eA_{e'}\right\} \\
&= 2\psi n^{-1}d(G) + 4n^{-2}\psi^2 2\sum\nolimits_{e<e'}A_eA_{e'}.
\end{aligned}$$

Meanwhile, average degree squared is $\left(\frac{2}{n}\sum_e A_e\right)^2 = 4n^{-2}\left\{\sum A_e + 2\sum_{e<e'}A_eA_{e'}\right\}$, which is useful since $\left(\mathrm{E}\left[v|G\right]\right)^2 = \left(\psi\frac{2}{n}\sum_e A_e\right)^2 = \psi^2 d(G)^2$. Therefore,

$$\mathrm{E}\left[v^2|G\right] = 2\psi n^{-1}d(G) + \psi^2\left\{4n^{-2}2\sum\nolimits_{e<e'}A_eA_{e'}\right\} = 2n^{-1}\psi\left(1-\psi\right)d(G) + \psi^2 d(G)^2.$$

It follows that $\mathrm{var}\left(v|G\right) = 2n^{-1}\psi\left(1-\psi\right)d(G) + \psi^2 d(G)^2 - \psi^2 d(G)^2 = 2n^{-1}\psi\left(1-\psi\right)d(G)$, completing the proof. ■

With the analytical formula for the variance of v, we can compute $\beta^{\star}$. Let $X := \psi \cdot (d(G_1), ..., d(G_R))'$ denote the scaled vector of true (unobserved) average degrees, $Z := (d(\bar{G}_1), ..., d(\bar{G}_R))$ the observed vector of sampled degrees, and $V := (v_1, ..., v_R) = Z - X$.

Then it is clear that $\beta^* := (Z'Z - V'V)^{-1} Z'y$ is consistent for β_0. By estimating $\Sigma_v :=$ plim $V'V/R$, we have $\beta^\star := \left(Z'Z - R\widehat{\Sigma}_v\right)^{-1} Z'y \xrightarrow{\mathrm{P}} \beta_0$. Under mild regularity conditions on the growth of average degree, we may therefore estimate $\beta^\star$. Estimation can be improved by performing a bootstrap bias correction.

H.2. **Numerical Evidence.** Table H.1 displays simulation results for network-level regression of average degree with simulated outcomes, precisely of the form presented in Figure 7. We choose ψ to make the expected edge count comparable to that of the induced subgraph. The results confirm the fact that the naive estimator exhibits significant biases, the analytical correction vastly reduces the biases but may still retain some residual bias which can further be mitigated by applying regularization.

Table H.1. Bias for Network-Level Regression for Average Degree (Simulated Graphs)

		Edge Sampling Rate				
Estimator for Regression on Avg. Degree	Property	(1/4)^2	(1/3)^2	(1/2)^2	(2/3)^2	1
β Naïve Estimator	% Bias	1361.2%	756.8%	291.4%	123.0%	0.0%
	Coverage	0%	0%	0%	0%	96%
β Analytic Correction	% Bias	-8.7%	-4.8%	-2.2%	-0.9%	0.0%
	Coverage	39%	61%	81%	87%	95%
β Regularized Analytic Correction	% Bias	0.3%	0.3%	-0.2%	-0.1%	0.0%
(Using Analytic Variance of Correction)	Coverage	97%	96%	96%	96%	95%
β Reg. Analytic Correction, Bootstrap BC	% Bias	-0.4%	-0.1%	-0.4%	-0.1%	0.0%
(Using Analytic Variance of Correction)	Coverage	97%	96%	96%	96%	95%

Estimators: The naïve estimator uses the subgraph generated by independently sampling edges at the given sampling rates. The analytic correction estimator, as noted, has residual bias. The regularized analytic correction adjusts for the dispersion term using an analytic formula for the variance of the remaining measurement error left after performing the analytic correction, which is only correct in expectation.

Notes: The sampling rates have been chosen to be comparable with the edge count in the $G^{|S}$ sampling simulations. Coverage computed using bootstrapped standard errors for all estimators. R2 was set to 0.7. There were 200 simulations done using the standard setup: homophilic networks with 6 groups, graphs with 250 nodes, and 50 graphs per regression.