

ESTIMATION OF PEER EFFECTS IN ENDOGENOUS SOCIAL NETWORKS: CONTROL FUNCTION APPROACH

Ida Johnsson and Hyungsik Roger Moon*

Abstract—We propose methods of estimating the linear-in-means model of peer effects in which the peer group, defined by a social network, is endogenous in the outcome equation for peer effects. Endogeneity is due to unobservable individual characteristics that influence both link formation in the network and the outcome of interest. We propose two estimators of the peer effect equation that control for the endogeneity of the social connections using a control function approach. We leave the functional form of the control function unspecified, estimate the model using a sieve semiparametric approach and establish asymptotics of the semiparametric estimator.

I. Introduction

THE ways in which interconnected individuals influence each other are usually referred to as peer effects. One of the first to formally model peer effects is Manski (1993), who proposes the linear-in-means model, in which an individual's action depends on the average action of other individuals and possibly also on their average characteristics. Manski assumes that all individuals within a given group are connected. Later literature allows for more complex patterns of connections, in which an individual might be directly influenced by a subset of the group. Examples are Bramoullé, Djebbari, and Fortin (2009), Lee, Liu, and Lin (2010), and Lee (2007a). Models of peer effects have been applied in education, health, and development, among others. Examples of applications are found in recent review papers such as Blume et al. (2011), Manski (2000), Epple and Romano (2011), Brock and Durlauf (2001), and Graham (2011).

Many models considered in earlier literature assume that connections between individuals are independent of unobserved individual characteristics that influence outcomes. However, assuming exogeneity of the network or peer group is restrictive in many applications. For example, consider the following widely studied empirical application of peer effects: peer influence on scholarly achievement. The assumption that friendships are exogenous in the outcome equation

for scholarly achievement means that there are no unobserved variables that influence both friendship formation and individual grades. However, even if a study controls for observable individual characteristics such as gender, age, race, and parents' education, it is likely to omit factors that influence both students' choice of friends and their GPA—for example, parental expectations, psychological disorders, or unreported substance use. For more examples of endogenous peer groups, see Brock and Durlauf (2001), Weinberg (2007), Shalizi (2012), and Hsieh and Lee (2016), among others.

In this paper we propose a method for estimating a linear-in-means model of peer effects, where the peer group is defined by a network that is endogenous in the outcome equation. Our model allows for correlation between the unobserved individual heterogeneity that has an impact on network formation and the unobserved characteristics of the outcome. For this, we use a dyadic network formation model that allows the unobserved individual attributes of two different agents to influence link formation and in which links are pairwise independent conditional on the observed and unobserved individual attributes. The network formation we consider in the paper is dense and nonparametric.

The main contributions of the paper are methodological. First, given the endogenous peer group formation, we show that we can identify the peer effects by controlling the unobserved individual heterogeneity of the network formation equation. Second, we propose an empirically tractable implementation of the control function, whose functional form is not parametrically specified. For this, we propose two approaches, one based on an estimator of the unobserved individual heterogeneity and the other one based on the average node degrees of the network.¹ Our estimation method is semiparametric because we do not restrict the functional form of the control function. Finally, we derive the limiting distributions of the estimators within a large single network. The main challenge of the asymptotics is handling the strong dependence of observables caused by the dense network. Other papers on peer effects that have considered endogenously formed peer groups and have controlled the endogeneity via various control functions include Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016), Qu and Lee (2015), Arduini et al. (2015), and Auerbach (2016). We provide more detail on these papers in section IIC.

The remainder of the paper is organized as follows. In section II, we present a high-level description of our approach and provide intuition as to its empirical applications. In section III, we formally present our model. In section IV, we show how to identify peer effects using control functions.

¹We acknowledge that this approach is developed based on an idea provided by one of the referees. We thank the referee.

Received for publication March 2, 2016. Revision accepted for publication August 14, 2019. Editor: Bryan S. Graham.

*Johnsson: University of Southern California; Moon: University of Southern California and Yonsei University.

We thank Bryan Graham and three referees for their helpful and valuable comments and suggestions. We are particularly grateful to one of the referees for suggesting the idea that is presented in section V.B. of the paper. We also appreciate the comments and discussions of the participants at the 2015 USC Dornsife INET Conference on Networks, the 2016 North American Summer Meeting of the Econometric Society, the 2016 California Econometrics Conference, the 2017 Asian Meeting of Econometric Society, the 2017 IAAE conference, the 2018 UCLA-USC Mini Conference, and the econometrics seminars at University of British Columbia and Ohio State University. The first draft of the paper was written while I.J. was a graduate fellow of USC Dornsife INET and H.R.M. was the associate director of USC Dornsife INET. H.R.M. acknowledges that this work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A2A01023679).

A supplemental appendix is available online at https://doi.org/10.1162/rest_a_00870.

Estimation is discussed in section V, and in section VI, we discuss the limiting distribution of the estimator and propose standard errors. In section VII, we present results of Monte Carlo simulations. There we compare the finite sample performance of our two semiparametric estimators against an estimator that assumes unobserved characteristics enter in a linear way, as well as an instrumental variables (IV) estimator that does not control for network endogeneity. We investigate both high-degree and low-degree networks. Section VIII concludes.

A word on notation: in what follows, we denote scalars by lowercase letters, vectors by lowercase bold letters, and matrices by uppercase bold letters.

II. Main Idea

In this section we introduce a simple model in order to illustrate the main points of our approach. A more general model and detailed discussion of the model follow later.

A. Simple Model

A simple peer effect model for the purpose of illustration of the main idea is

$$y_i = \beta^0 \left(\frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}} \right) + v_i, \quad i = 1, \dots, N, \quad (1)$$

where x_i is a measure of observable characteristics of individual i and d_{ij} is an indicator of individual i 's peer, so $d_{ij} = 1$ if i and j are directly linked and 0 otherwise. In equation (1), the regressor of interest is the average of the characteristics of those individuals who are linked with i , $\frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}}$. For simplicity, we assume that x_i is exogenous with respect to all the unobserved components of the model; this will be relaxed later.

For the link formation, we consider the following dyadic network formation model,

$$d_{ij} = \mathbb{I}(g(a_i, a_j) \geq u_{ij}) \mathbb{I}(i \neq j), \quad (2)$$

where a_i and a_j are unobserved individual specific characteristics, u_{ij} is a link-specific component, and $g(\cdot, \cdot)$ is some function. It should be noted that this model of network formation does not allow for network effects in link formation, as a link between i and j depends on only the characteristics of i and j .

The unobserved individual characteristic a_i can be interpreted as social capital that increases the likelihood of forming a link. Depending on the context, this could be factors like trustworthiness, socioeconomic status, or outspokenness.

For example, De Weerd and Fafchamps (2011) measure the risk-sharing links between households in Tanzania and construct links between households based on asking whom individuals could "personally rely on for help." Fafchamps

and Gubert (2007) examine the formation of risk-sharing networks using data from the rural Philippines. Banerjee et al. (2013) examine how participation in microfinance diffuses through a social network that they measure using lending and trust. In these settings, we can think of a_i as a measure of individual trustworthiness and integrity in financial matters. Ductor et al. (2014) analyze whether knowledge of a researcher's coauthorship network is helpful in predicting his or her productivity. In this setting, a_i can be interpreted as some unobserved productivity trait that induces the researcher to have more coauthors and also to be more productive at writing papers.

B. Control Function and Its Implementation

The key feature of the peer effect model, equations (1) and (2), is that individual i 's unobserved characteristic a_i , which affects link formation, is correlated with v_i , i 's unobserved characteristic that affects the outcome y_i . For example, a_i could be an unobserved component that affects a researcher's publication rate y_i and also his or her coauthorship relationships, d_{ij} . Alternatively, we can think of a situation where there are two types of agents: popular and unpopular. The popular agents are more likely to be friends with other agents, and popular agents have better outcomes even in the absence of a peer effect. Then the peer formation d_{ij} becomes correlated with the unobserved component v_i of the outcome, and, as a consequence, the regressor of the peer effect, $\frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}}$, becomes endogenous.

In this paper we use a control function method to handle the endogenous peer group problem. Let \mathbf{D}_N be the $N \times N$ adjacency matrix that describes the network links d_{ij} . Suppose that the unobserved characteristics (a_i, v_i) and u_{ij} are randomly drawn over i and (i, j) , respectively. Also assume that u_{ij} is independent of (a_i, v_i) . Then, for any $i \neq j$, the link $d_{ij} = \mathbb{I}(g(a_i, a_j) \geq u_{ij})$ and v_i are dependent only through a_i . Therefore, controlling for a_i , the network \mathbf{D}_N and v_i become mean independent, that is,

$$\mathbb{E}(v_i | \mathbf{D}_N, a_i) = \mathbb{E}(v_i | a_i) =: h(a_i).$$

Suppose that we observe a_i . Consider the outcome equation that controls for a_i nonparametrically,

$$y_i = \beta^0 \left(\frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}} \right) + h(a_i) + \varepsilon_i,$$

where $\varepsilon_i := v_i - h(a_i)$. Once we control the endogeneity of the network with a_i , the regressor of the peer effect becomes exogenous, and we can estimate the peer effect coefficient β^0 using the conventional partially linear regression estimation method (Robinson, 1988).

However, in most empirical applications, a_i is not observed. Then the question becomes how to implement the control function. In this paper, as the main methodological

contribution, we propose the following two procedures, both implemented with a single snapshot of an observed network.

First, suppose that a_i can be consistently estimated. An example can be found in Graham (2017) with the specification $g(a_i, a_j) = a_i + a_j$. Then we estimate β^0 by running the partially linear regression of y_i on $\frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}}$ and $h(\hat{a}_i)$ as in Robinson (1988).

The second method is to use an observed control function that asymptotically carries the same information as a_i . For this, first notice by the WLLN,

$$\begin{aligned} \deg_i &:= \frac{1}{N} \sum_{j \neq i} d_{ij} \\ &= \frac{1}{N} \sum_{j \neq i} \mathbb{I}(g(a_i, a_j) \geq u_{ij}) \rightarrow_p \mathbb{P}(d_{ij} = 1 | a_i). \end{aligned}$$

Suppose that the network formation probability conditional on a_i , $\mathbb{P}(d_{ij} = 1 | a_i)$ is a monotonic function of a_i . A sufficient condition for this is that $g(\cdot, a_j)$ is monotonic in the same direction for all a_j —for example,

$$g(a_i, a_j) = a_i + a_j - \tau |a_i - a_j|, \quad (3)$$

with $0 \leq \tau < 1$. In this case, the limit of the average node degree, $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j \neq i} d_{ij}$, carries the same information as the control function a_i , which justifies \deg_i as a proxy of the control function a_i , that is, $\mathbb{E}(v_i | a_i) \simeq \mathbb{E}(v_i | \deg_i) =: h_*(\deg_i)$. The peer effect coefficient β^0 can be estimated by using \deg_i as a control function. More specifically, we estimate β^0 by running the partially linear regression of y_i on $\frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}}$ and $h_*(\deg_i)$. Intuitively, unobserved characteristics a_i drive heterogeneous degree sequences. We can therefore control for degree when estimating peer effects, ignoring the specific choice of a structural model explaining heterogeneous degrees.

The use of degree as a control function requires many fewer restrictions on the specification of the network. Intuitively, the unobserved node (or individual) fixed effects a_i control for heterogeneous degree sequences. Therefore, from an economic point of view, what needs to be controlled is the agent's degree, which validates the control function approach that uses \deg_i . This approach does not require a specification of the specific structural model explaining heterogeneous degree sequences. Consistent estimation of a_i usually requires a specific functional form. For example, Graham (2017) assumed an additive model, and Chen, Fernández-Val, and Weidner (2014) require an interactive form. However, there is a disadvantage in the degree approach: it cannot identify the coefficient of the observed exogenous regressor if the same regressor also impacts the network formation.

In section III, we generalize the simple model, equation (1), by allowing for an additional peer effect, $\frac{\sum_{j \neq i} d_{ij} y_j}{\sum_{j \neq i} d_{ij}}$, known as

the endogenous peer effect, which measures the effects of the outcomes of the peer group on an individual outcome. In this case, we have to deal with two kinds of endogeneity in the peer effect regressors: one from the endogenous regressors y_j and the other from the endogenous peers d_{ij} . In section III, we also generalize the dyadic network formation model by introducing a dyadic component based on observed individual characteristics. We provide application examples of the general model and discuss its features there. The identification of the peer effects in the general model will be discussed in section IV. In section V, we show how to implement the two estimation methods in the general framework. In the appendix, we provide the regularity conditions that are required for the asymptotic results of the paper. All the technical proofs and comprehensive Monte Carlo simulation results are found in the online supplement, available at https://doi.org/10.1162/rest_a_00870.

C. Related Literature

Closely related papers that adopt a control function approach include Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016), Qu and Lee (2015), Arduini et al. (2015), and Auerbach (2016). Our paper adopts a frequentist approach based on a nonparametric specification of the network formation, while Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016) use the Bayesian method based on a full parametric specification of the network formation and the outcome equation. Like our paper, Qu and Lee (2015) assume the network (spatial weights in their model) to be endogenous through unobserved individual heterogeneity. However, our paper is different from Qu and Lee (2015) in many ways. They consider sparse network formation models, while we consider a dense network. They restrict the functional form of the control function to be linear, while we impose no restriction on the functional form. The two papers propose different implementations of the control function. Also, in Goldsmith-Pinkham and Imbens (2013), unobserved components account for homophily in link formation, whereas in our setup, they mainly drive degree heterogeneity but are allowed to account for homophily as well, as in example (3).

Our paper is different from Arduini et al. (2015) regarding the main source of the endogeneity of the network and the form of the control function. They assume that the endogeneity of the network is allowed through dependence between the outcome equation error and the idiosyncratic network formation error, like the conventional sample selection model. This model can be interpreted as meeting opportunities being correlated with unobserved ability of the agent that affects the outcome. They also consider control functions (both parametric and semiparametric) to deal with the selection bias problem and propose a semiparametric estimator that uses a power series to approximate selectivity bias terms. Both Qu and Lee (2015) and Arduini et al. (2015) derive the asymptotics using near-epoch dependence and are based on the

assumption that the number of connections does not increase at the same rate as the square of the network size.

Among the related papers, probably the one most closely related to ours is Auerbach (2016). As a result, we discuss the differences between the two papers in more detail. The outcome model of Auerbach (2016) is a partially linear regression model where the nonparametric component is an unknown function of the unobserved network heterogeneity,

$$y_i = \beta^0 x_i + h(a_i) + \varepsilon_i, \\ d_{ij} = \mathbb{I}(g(a_i, a_j) \geq u_{ij}) \mathbb{I}(i \neq j).$$

In the simple peer effect example, the exogenous peer effect corresponds to the regressor x_i above. The network formation is the same as equation (2).

To compare the identification ideas, let's assume that $a_i \sim U[-1/2, 1/2]$ and $u_{ij} \sim U[0, 1]$. In this case, $d_i := (d_{i1}, \dots, d_{in})'$ and the distribution of d_i of node i , whose characteristic is a_i , is fully characterized by the link formation probability profile $g(a_i, \bullet)$.

The key condition of Auerbach (2016) is that $h(a_i)$ and the link formation distribution profile $g_i(\bullet) := g(a_i, \bullet)$ be one-to-one a.s., that is, $g(a, \bullet) \neq g(a^*, \bullet)$ a.s., if and only if $h(a) \neq h(a^*)$. Then, for any distance measure between the two profiles g_i and g_j , $d(g_i, g_j)$, it follows that $d(g_i, g_j) = 0$ if and only if $h(a_i) = h(a_j)$.

Based on this, Auerbach (2016) finds that one can control network endogeneity by pair-wise differencing² of the observations of the two individuals, i and j , whose network formation distributions are the same, $d(g_i, g_j) = 0$, and proposes a semiparametric estimator based on matching pairs of agents with similar columns of the squared adjacency matrix.

Notice that the identification condition of Auerbach (2016) is satisfied if $g(a_i, \bullet)$ and a_i have a one-to-one relation. However, our second identification is based on the condition that a_i and the marginal network probability, $\int g(a_i, \tau) d\tau$, have a one-to-one relation. We admit that this condition is more restrictive than the identification condition of Auerbach (2016) because our restriction is a special case of his restriction. However, our identification under the stronger condition allows for the omitted variable in the peer effects equation to be nonparametrically directly estimated, which results in the peer effect estimator having the parametric convergence rate (\sqrt{N}). This feature is not necessarily guaranteed in the framework of Auerbach (2016).³

III. General Model of Peer Effects with an Endogenous Network

In this section, we introduce a general linear-in-means peer effect model that extends the simple illustrative out-

come model with a peer effect in equation (1) and the simple dyadic network formation model in equation (2).

A. General Linear-in-Means Peer Effects Model

As in section II, d_{ij} are the observed binary variables that measure undirected links among individuals $i \in \{1, 2, \dots, N\}$. We assume that individual outcomes are given by the linear-in-means model of peer effects

$$y_i = \left(\sum_{j=1, j \neq i}^N g_{ij} y_j \right) \beta_1^0 + \mathbf{x}_{1i}' \beta_2^0 + \left(\sum_{j=1, j \neq i}^N g_{ij} \mathbf{x}_{1j} \right)' \beta_3^0 + v_i, \quad (4)$$

where \mathbf{x}_{1i} are observed individual characteristics that affect the outcome y_i , v_i are unobserved individual characteristics, and

$$g_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{d_{ij}}{\sum_{j \neq i} d_{ij}} & \text{otherwise} \end{cases}$$

is the weight of the peer effects. Using the terminology of Manski (1993), β_1^0 captures the endogenous social effect, and β_3^0 measures the exogenous social effect. We let $\beta^0 := (\beta_1^0, \beta_2^0, \beta_3^0)'$ and denote $\beta = (\beta_1, \beta_2, \beta_3)'$.

We let \mathbf{D}_N be the $(N \times N)$ adjacency matrix of the network whose (i, j) th element is d_{ij} . We let $d_{ii} = 0$ for all i , following convention. Let \mathbf{G}_N be the matrix whose (i, j) th element is g_{ij} . Recall that \mathbf{G}_N is obtained by row-normalizing \mathbf{D}_N . Denote $\mathbf{X}_{1N} = (\mathbf{x}_{11}', \dots, \mathbf{x}_{1N}')'$, $\mathbf{y}_N = (y_1, \dots, y_N)'$, and $\mathbf{v}_N = (v_1, \dots, v_N)'$. Using this notation, we can express the linear-in-means peer effects model, equation (4), as

$$\mathbf{y}_N = \mathbf{G}_N \mathbf{y}_N \beta_1^0 + \mathbf{X}_{1N} \beta_2^0 + \mathbf{G}_N \mathbf{X}_{1N} \beta_3^0 + \mathbf{v}_N. \quad (5)$$

Throughout the paper, we assume that $|\beta_1^0| < 1$. It is known that when \mathbf{G}_N is row-normalized (i.e., $\sum_{j \neq i} g_{ij} = 1$) and $|\beta_1^0| < 1$, the (equilibrium) solution of the peer effect model uniquely exists (e.g., see Bramoullé et al., 2009) as

$$\mathbf{y}_N = (\mathbf{I}_N - \beta_1^0 \mathbf{G}_N)^{-1} (\mathbf{X}_{1N} \beta_2^0 + \mathbf{G}_N \mathbf{X}_{1N} \beta_3^0 + \mathbf{v}_N) \\ = \sum_{k=0}^{\infty} (\beta_1^0 \mathbf{G}_N)^k (\mathbf{X}_{1N} \beta_2^0 + \mathbf{G}_N \mathbf{X}_{1N} \beta_3^0 + \mathbf{v}_N). \quad (6)$$

In the standard linear-in-means model of peer effects, the main focus has been identification and estimation of peer effects, assuming that the peer group (or the network) is exogenous, that is, $\mathbb{E}[v_i | \mathbf{X}_{1N}, \mathbf{G}_N] = 0$ —for example, see Manski (1993), Bramoullé et al. (2009), Lee (2007a), and Blume et al. (2015). To identify and estimate the linear-in-means model of peer effects when the peer group is exogenous, it is necessary to take into account the fact that the regressor $\sum_{i=1}^N g_{ij} y_j$ is correlated with the error term v_i . For example,

²This resembles Powell (1987), Heckman, Ichimura, and Todd (1998), and Abadie and Imbens (2006).

³We thank one of the referees for suggesting the comparisons.

if $v_i \sim \text{i.i.d.}(0, \sigma^2)$, it is true that

$$\begin{aligned} \mathbb{E}[(\mathbf{G}_N \mathbf{y}_N)' \mathbf{v}_N] &= [(\mathbf{G}_N(\mathbf{I}_N - \beta_1^0 \mathbf{G}_N)^{-1}(\mathbf{X}_{1N} \beta_2^0 \\ &\quad + \mathbf{G}_N \mathbf{X}_{1N} \beta_3^0 + \mathbf{v}_N))' \mathbf{v}_N] \\ &= \mathbb{E}[(\mathbf{G}_N(\mathbf{I}_N - \beta_1^0 \mathbf{G}_N)^{-1} \mathbf{v}_N)' \mathbf{v}_N] \\ &= \sigma_0 \text{tr}(\mathbf{G}_N(\mathbf{I}_N - \beta_1^0 \mathbf{G}_N)^{-1}) \neq 0. \end{aligned} \quad (7)$$

To solve this endogeneity problem different estimators have been proposed in the literature, for example, in Kelejian and Prucha (1998) and Lee (2003, 2007b). One of the widely used estimation methods is the IV approach. In view of the expression of equation (6), when $\beta_2^0 \neq 0$, we can use $\mathbf{G}_N^2 \mathbf{X}_{1N}$ as the IV of the endogenous regressor $\mathbf{G}_N \mathbf{y}_N$ because $\mathbf{G}_N^2 \mathbf{X}_{1N}$ is uncorrelated with \mathbf{v}_N while it is correlated with the endogenous regressor $\mathbf{G}_N \mathbf{y}_N$ (see Kelejian & Prucha, 1998; Lee, 2003; and Bramoullé et al., 2009).⁴ Then, the natural estimator is the two-stage least squares (2SLS) estimator,

$$\hat{\beta}_N^{2SLS} = (\mathbf{W}_N' \mathbf{Z}_N (\mathbf{Z}_N' \mathbf{Z}_N)^{-1} \mathbf{Z}_N' \mathbf{W}_N)^{-1} \mathbf{W}_N' \mathbf{Z}_N (\mathbf{Z}_N' \mathbf{Z}_N)^{-1} \mathbf{Z}_N' \mathbf{y}_N, \quad (8)$$

where $\mathbf{W}_N = [\mathbf{G}_N \mathbf{y}_N, \mathbf{X}_{1N}, \mathbf{G}_N \mathbf{X}_{1N}]$ and $\mathbf{Z}_N = [\mathbf{X}_{1N}, \mathbf{G}_N \mathbf{X}_{1N}, \mathbf{G}_N^2 \mathbf{X}_{1N}]$ is the matrix of instruments. For the IVs \mathbf{Z}_N to be strong, we assume that $\beta_2^0 \neq 0$.

When the network matrix is endogenous, $\mathbb{E}[\mathbf{G}_N \mathbf{v}_N] \neq 0$, and the procedure used by Kelejian and Prucha (1998), Lee (2003), Bramoullé et al. (2009), and others is no longer valid since the IV matrix $\mathbf{Z}_N = [\mathbf{X}_{1N}, \mathbf{G}_N \mathbf{X}_{1N}, \mathbf{G}_N^2 \mathbf{X}_{1N}]$ is correlated with the error term \mathbf{v}_N . Specifically, the validity of the 2SLS estimator depends on the orthogonality condition $\mathbb{E}[\mathbf{v}_N | \mathbf{Z}_N] = 0$, which is implied if $\mathbb{E}[\mathbf{v}_N | \mathbf{X}_{1N}, \mathbf{G}_N] = 0$. However, it does not hold if the (row-normalized) network \mathbf{G}_N is correlated with \mathbf{v}_N , which is true if unobserved individual characteristics of \mathbf{G}_N directly influence both link formation and individual outcomes.

In this paper, we consider the case where it may be that $\mathbb{E}[\mathbf{v}_N | \mathbf{X}_{1N}, \mathbf{G}_N] \neq 0$, so that unobserved characteristics that influence link formation can also have a direct effect on individual outcomes. This is an important consideration in many common applications, like the impact of school friendships on scholarly achievement or substance use. Imagine kids from homes where parents help with homework who only form friendships with kids from similar homes. If this unobserved characteristic of parental behavior is not taken into account and if this is what really determines grades, this effect might falsely be classified as a peer effect. A more elaborate discussion of our framework and its empirical applications is in section II.

⁴If $\beta_2^0 = 0$, \mathbf{y}_N does not depend on \mathbf{X}_{1N} and $\mathbf{G}_N^2 \mathbf{X}_{1N}$ is not a relevant instrument for $\mathbf{G}_N \mathbf{y}_N$.

B. Model of Network Formation

Let \mathbf{x}_{2i} be a vector of observable characteristics of individual i , and let $\mathbf{x}_i = \mathbf{x}_{1i} \cup \mathbf{x}_{2i}$. Define \mathbf{X}_{2N} analogous to \mathbf{X}_{1N} and let $\mathbf{X}_N = \mathbf{X}_{1N} \cup \mathbf{X}_{2N}$. We introduce a_i , a scalar unobserved characteristic of individual i , which is treated as an individual fixed effect and, hence, might be correlated with \mathbf{x}_i . We denote the vector of individual unobserved characteristics by $\mathbf{a}_N = (a_1, a_2, \dots, a_N)'$. Individuals are connected by an undirected network \mathbf{D}_N , with the (i, j) th element $d_{ij} = 1$ if i and j are directly connected and 0 otherwise. We assume the network to be undirected,⁵ $d_{ij} = d_{ji}$, and assume $d_{ii} = 0$ for all i , following the convention. In this case, there are $n = \binom{N}{2}$ dyads. Let \mathbf{t}_{ij} denote an $l_T \times 1$ vector of dyad-specific characteristics of dyad ij , and we assume that $\mathbf{t}_{ij} = t(\mathbf{x}_{2i}, \mathbf{x}_{2j})$. Agents form links according to

$$d_{ij} = \mathbb{I}(g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) - u_{ij} \geq 0), \quad (9)$$

where $\mathbb{I}(\bullet)$ is an indicator function. In this setup, link surplus is transferable across directly linked agents and consists of three components: $\mathbf{t}_{ij} := t(\mathbf{x}_{2i}, \mathbf{x}_{2j})$ is a systematic component that varies with observed dyad attributes and accounts for homophily, a_i and a_j account for unobserved dyad attributes (degree heterogeneity), and u_{ij} is an idiosyncratic shock that is i.i.d. across dyads and independent of \mathbf{t}_{ij} and a_i for all i, j . Since links are undirected, the surplus of link d_{ij} must be the same for individuals i and j . Hence, we assume that the function t_{ij} is symmetric in i and j , and the function g is symmetric in a_i and a_j .

In the literature, various parametric versions of the network formation in equation (9) are used (Jackson, 2005; Graham, 2017). An important example of a parametric specification is the one in Graham (2017):

$$d_{ij} = \mathbb{I}(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})' \lambda + a_i + a_j - u_{ij} > 0). \quad (10)$$

For the purpose of the paper, particularly in constructing the estimators that we introduce in section V, we do not need a parametric specification.

Regarding the network formation, equation (9), we impose restrictions (assumption 11, iii–vi, in the appendix) that imply the following two features. The first feature is that the link formation probability of individual i with characteristics (\mathbf{x}_{2i}, a_i) is one-to-one with respect to the unobserved characteristic a_i , that is, for all \mathbf{x}_{2i} ,

$$a_i \neq a_i^* \text{ if and only if } \mathbb{P}(d_{ij} = 1 | \mathbf{x}_{2i}, a_i) \neq \mathbb{P}(d_{ij} = 1 | \mathbf{x}_{2i}, a_i^*). \quad (11)$$

Obviously, this condition is satisfied in the parametric model, equation (10). This monotonic condition justifies the use of the average node degree in implementing the control function

⁵Our analysis can be extended to the directed network case, but we do not pursue it in this paper.

as introduced in section II and will be discussed in section VB. The second feature is that the network formed by equation (9) is dense in the sense that the expected number of connections is proportional to the square of the network size. This is satisfied if the error u_{ij} is drawn randomly from a distribution with full support, while $g(\mathbf{t}_{ij}, a_i, a_j)$ is bounded (see assumption 11 iii–v in the appendix). In this case, the probability of any two individuals forming a link is bounded away from 0 and strictly less than 1. The dense network model is appropriate for scenarios where any two individuals can plausibly form a link. Notice that the dense network assumption and the sharing restriction on the net surplus function g are necessary for implementing the control function in section V and establishing the asymptotic theory of the control function based estimators in section VI. If a_i is observed, we can identify and estimate peer effects without these assumptions (see section IV).

Regarding the network formation model, equation (9), it is important to note that this model rules out interdependent link preferences, and it assumes that links are formed independently conditional on observed individual characteristics and unobserved fixed effects. As discussed in Graham (2017), this assumption is appropriate for settings where link formation is driven predominantly by bilateral concerns, such as certain types of friendship networks and trade networks and some models of conflict between nation-states. The model in equation (9) is not a good choice when important strategic aspects influence link formation, like when the identity of the nodes to which j is linked influences i 's return from forming a link with j . A discussion of networks with interdependent links can be found in Graham (2017) and De Paula (2017). Also, when network externalities are present, the additional complication of multiple equilibria has to be considered (see Sheng, 2012, for more details).

IV. Identification of Peer Effects Using a Control Function Approach

In this section we provide an identification argument for the peer effect equation based on a control function when the network is endogenous.

A. Control Function of Network Endogeneity

In this section we discuss how to control the endogeneity of the peer group defined by the network formed in equation (9). First, we introduce a basic assumption that we will maintain throughout the paper:

Assumption 1. (i) (\mathbf{x}_i, a_i, v_i) are i.i.d. for all $i, i = 1, \dots, N$, (ii) $\{u_{ij}\}_{i,j=1,\dots,N}$ are independent of $(\mathbf{X}_N, \mathbf{a}_N, \mathbf{v}_N)$ and i.i.d. across (i, j) with cdf $\Phi(\cdot)$, and (iii) $\mathbb{E}(v_i|\mathbf{x}_i, a_i) = \mathbb{E}(v_i|a_i)$.

Assumption 1(i) implies that the observables \mathbf{x}_i and the unobservable characteristics (a_i, v_i) are randomly drawn. This is a standard assumption in the peer effects literature. Assumption 1(ii) assumes that the link formation error u_{ij} is

orthogonal to all other observables and unobservables in the model. This means that the dyad-specific unobservable shock u_{ij} from the link formation process does not influence outcomes $(y_1, \dots, y_N)'$. However, we allow for endogeneity of the social interaction group through dependence between the two unobserved components a_i and v_i . This means that the unobserved error v_i in the outcome equation can be correlated with unobserved individual characteristics a_i that are determinants of link formation. We also allow the observed characteristics \mathbf{x}_i of the outcome equation and the network formation to be correlated with the unobserved components (v_i, a_i) , so that the regressor \mathbf{x}_{1i} can be endogenous in the outcome equation, and the network formation observables \mathbf{x}_{2i} can be arbitrarily correlated with the unobserved individual characteristic a_i . In assumption 1(iii), we assume that the dependence between \mathbf{x}_i and v_i exists only through a_i . That is, a_i is the fixed effect of individual i and controls the endogeneity of \mathbf{x}_i with respect to v_i .

Notice that the network \mathbf{D}_N defined in equation (9) and the (row normalized) network \mathbf{G}_N are measurable functions of $(\mathbf{x}_{2i}, \mathbf{x}_{2,-i}, a_i, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\dots,N})$, where $\mathbf{x}_{2,-i} = (\mathbf{x}_{2,1}, \dots, \mathbf{x}_{2,i-1}, \mathbf{x}_{2,i+1}, \dots, \mathbf{x}_{2,N})$ and \mathbf{a}_{-i} is defined analogously. Under assumption 1, we have

$$\begin{aligned} \mathbb{E}[v_i|\mathbf{X}_N, \mathbf{G}_N, a_i] &= \mathbb{E}[v_i|\mathbf{x}_{-i}, \mathbf{G}_N(\mathbf{x}_{2,-i}, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\dots,N}, \mathbf{x}_{2i}, a_i), \mathbf{x}_i, a_i] \\ &= \mathbb{E}[v_i|\mathbf{x}_i, a_i] = \mathbb{E}[v_i|a_i], \end{aligned} \quad (12)$$

where the second equality holds because $(\mathbf{x}_{-i}, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\dots,N})$ and (\mathbf{x}_i, a_i, v_i) are independent under assumptions 1(i) and 1(ii). This shows v_i and $(\mathbf{x}_{-i}, \mathbf{G}_N(\mathbf{x}_{2,-i}, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\dots,N}, \mathbf{x}_{2i}, a_i))$ are mean-independent conditioning on (\mathbf{x}_i, a_i) . The last line follows by the fixed effect assumption, assumption 1(iii).

The result, equation (12), shows that conditional on the unobserved heterogeneity a_i in the network formation (and any subcomponents of \mathbf{x}_i), the unobserved characteristic v_i that affects the outcome y_i becomes uncorrelated with the (row-normalized) network \mathbf{G}_N (and the observables \mathbf{X}_N). This implies that the network endogeneity can be controlled by a_i (or together with any subcomponents of \mathbf{x}_i). We summarize the discussion in the following lemma:

Lemma 1 (Control Function of Peer Group Endogeneity). *Suppose that assumption 1 holds. Then, $\mathbb{E}[v_i|\mathbf{X}_N, \mathbf{G}_N, a_i] = \mathbb{E}[v_i|\mathbf{x}_i, a_i]$.*

B. Identification of Peer Effects with a_i as Control Function

In this section we show how to identify the peer effects in the outcome question when the endogenous network is formed by equation (9). We provide two identification methods depending on whether we control the network (peer group) endogeneity with a_i or a_i together with \mathbf{x}_{2i} , in the case when \mathbf{x}_{2i} and \mathbf{x}_{1i} do not overlap.

First, notice that regardless of the possible endogeneity of the (row-normalized) network \mathbf{G}_N , we need to control for the endogeneity of the term $\sum_{j \neq i} g_{ij} y_j$ that represents the so-called endogenous peer effects. When the peer group \mathbf{G}_N is exogenous and uncorrelated with v_N , $\mathbf{G}_N^2 \mathbf{X}_{1N}$ is often used as an IV for the endogenous peer effects term $\mathbf{G}_N \mathbf{y}_N$ (Kelejian & Prucha, 1998; Lee, 2003; and Bramoullé et al., 2009).

Let $\mathbf{Z}_N = [\mathbf{X}_{1N}, \mathbf{G}_N \mathbf{X}_{1N}, \mathbf{G}_N^2 \mathbf{X}_{1N}]$ be the usual IV matrix used in 2SLS estimation of the peer effects equation. Note that \mathbf{Z}_N is no longer a valid IV matrix in our framework because the peer group defined by the network \mathbf{G}_N is correlated with v_N due to potential correlation between the unobserved v_i and a_i . Let $\mathbf{W}_N = [\mathbf{G}_N \mathbf{y}_N, \mathbf{X}_{1N}, \mathbf{G}_N \mathbf{X}_{1N}]$. Further, denote the transpose of the i th row of \mathbf{Z}_N and \mathbf{W}_N by \mathbf{z}_i and \mathbf{w}_i , respectively.

Suppose that assumption 1 holds, and so a_i controls the network endogeneity. Then,

$$\begin{aligned} & \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i])(v_i - \mathbb{E}(v_i | a_i)) | a_i] \\ &= \mathbb{E}[\mathbf{z}_i v_i | a_i] - \mathbb{E}[\mathbf{z}_i | a_i] \mathbb{E}[v_i | a_i] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{z}_i v_i | a_i, \mathbf{X}_{1N}, \mathbf{G}_N] | a_i] - \mathbb{E}[\mathbf{z}_i | a_i] \mathbb{E}[v_i | a_i] \\ &= \mathbb{E}[\mathbf{z}_i \mathbb{E}[v_i | a_i, \mathbf{X}_{1N}, \mathbf{G}_N] | a_i] - \mathbb{E}[\mathbf{z}_i | a_i] \mathbb{E}[v_i | a_i] \\ &\stackrel{(1)}{=} \mathbb{E}[\mathbf{z}_i \mathbb{E}[v_i | a_i] | a_i] - \mathbb{E}[\mathbf{z}_i | a_i] \mathbb{E}[v_i | a_i] \\ &= 0, \end{aligned} \quad (13)$$

where equality (1) holds by lemma 1(i). This shows that the instrumental variables \mathbf{z}_i or $\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i]$ become orthogonal to $v_i - \mathbb{E}[v_i | a_i]$, the residual of v_i after projecting out a_i .

Furthermore, if $\mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i | a_i])']$ has full rank, then we can identify the peer effect coefficients β^0 as

$$\begin{aligned} 0 &= \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i])(y_i - \mathbf{w}_i' \beta - \mathbb{E}[y_i - \mathbf{w}_i' \beta | a_i])] \\ &= \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i | a_i])'(\beta - \beta^0) \\ &\quad + \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i])(v_i - \mathbb{E}[v_i | a_i])] \\ &\stackrel{(1)}{=} \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i | a_i])'(\beta - \beta^0) \\ &\stackrel{(2)}{\Leftrightarrow} \beta = \beta^0, \end{aligned}$$

where equality (1) follows by the orthogonality result in equation (13) and equality (2) follows from the full rank condition.

Assumption 2 (Rank Condition). $\mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i | a_i])']$ has full rank.

For the full rank condition in assumption 2, it is necessary that the IVs \mathbf{z}_i and the regressors \mathbf{w}_i have additional variation after projecting out the control function a_i . As shown in the supplementary appendix A.2.3, when N is large, both \mathbf{z}_i and \mathbf{w}_i become close to functions that depend only on (\mathbf{x}_i, a_i) . In this case, for the full rank condition to be satisfied, it is necessary that there be additional random components in \mathbf{x}_i

that are different from a_i , so that the limits of \mathbf{z}_i and \mathbf{w}_i are not linearly dependent. As a summary, we have the following first identification theorem.

Theorem 1 (Identification). Under assumptions 1 and 2, the parameter β^0 is identified by the moment condition $\mathbb{E}[(\mathbf{z}_i - \mathbb{E}(\mathbf{z}_i | a_i))(y_i - \mathbb{E}(y_i | a_i) - (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i | a_i))' \beta^0)] = 0$:

$$\begin{aligned} & \mathbb{E}[(\mathbf{z}_i - \mathbb{E}(\mathbf{z}_i | a_i))(y_i - \mathbb{E}(y_i | a_i) - (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i | a_i))' \beta)] = 0 \\ & \Leftrightarrow \beta = \beta^0. \end{aligned}$$

Theorem 1 shows that we can identify the parameter β^0 by controlling the unobserved network heterogeneity a_i in the outcome equation and taking the residuals $y_i - \mathbb{E}(y_i | a_i) - (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i | a_i))' \beta$ and using the instrumental variables $\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | a_i]$.

C. Identification of Peer Effects Using (\mathbf{x}_{2i}, a_i) as Control Function

In view of the derivation of the control function in equation (12) under assumption 1, it is possible to use any regressors in \mathbf{x}_i in addition to the unobserved heterogeneity a_i . In this section, we discuss identification of the peer effects using (\mathbf{x}_{2i}, a_i) as a control function. The reason to consider this particular control function is that we can implement it in the absence of a consistent estimator of a_i , which we discuss in detail in section V.

First, suppose that there is no overlap between the regressors in the outcome equation \mathbf{x}_{1i} and the regressors in the network formation equation \mathbf{x}_{2i} and assume the conditions in assumption 1.⁶

Assumption 3. Assume that the conditions i to iii of assumption 1 hold. Also, assume that (iv) the explanatory variables in \mathbf{x}_{1i} and \mathbf{x}_{2i} do not overlap (i.e., $\mathbf{x}_{1i} \cap \mathbf{x}_{2i} = \emptyset$).

Then, under assumption 1 and by equation (12), it follows that

$$\mathbb{E}[v_i | \mathbf{X}_N, \mathbf{G}_N, a_i] = \mathbb{E}[v_i | a_i] = \mathbb{E}[v_i | \mathbf{x}_{2i}, a_i], \quad (14)$$

where the last line holds by assumption 1(iii). Then, similar to equation (13), we can show that

$$\mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | \mathbf{x}_{2i}, a_i])(v_i - \mathbb{E}(v_i | \mathbf{x}_{2i}, a_i)) | \mathbf{x}_{2i}, a_i] = 0. \quad (15)$$

Furthermore, suppose that the following full rank assumption is satisfied:

Assumption 4 (Rank Condition). $\mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i | \mathbf{x}_{2i}, a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i | \mathbf{x}_{2i}, a_i])']$ has full rank.

⁶Later in this section, we will discuss a more general case where \mathbf{x}_{1i} and \mathbf{x}_{2i} intersect.

Notice that if \mathbf{x}_{1i} and \mathbf{x}_{2i} are overlapped, then the full rank condition in assumption 4 does not hold.

Using similar arguments that lead to theorem 1, we can identify the peer effect coefficients β^0 as

$$0 = \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])(y_i - \mathbf{w}'_i\beta - \mathbb{E}[y_i - \mathbf{w}'_i\beta|\mathbf{x}_{2i}, a_i])] \\ \Leftrightarrow \beta = \beta^0. \quad (16)$$

This is summarized in the following theorem.

Theorem 2 (Alternative Identification). *Under assumptions 1, 3, and 4, the parameter β^0 is identified by the moment condition:*

$$\mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])(y_i - \mathbb{E}[y_i|\mathbf{x}_{2i}, a_i] - (\mathbf{w}'_i - \mathbb{E}[\mathbf{w}_i|\mathbf{x}_{2i}, a_i])'\beta) = 0 \iff \beta = \beta^0.$$

So far, we have considered the case where the regressors \mathbf{x}_{1i} and \mathbf{x}_{2i} do not intersect. A more general case is when the regressors \mathbf{x}_{1i} consist of two components, where one component is different from the observed control function \mathbf{x}_{2i} and the other is part of \mathbf{x}_{2i} . That is, $\mathbf{x}_{1i} = (\mathbf{x}_{11i}, \mathbf{x}_{12i})$, where \mathbf{x}_{11i} does not share any elements with \mathbf{x}_{2i} and \mathbf{x}_{11i} is nonempty, and $\mathbf{x}_{12i} \subset \mathbf{x}_{2i}$. Let $\beta_2^0 = (\beta_{21}^0, \beta_{22}^0)$, $\beta_3^0 = (\beta_{31}^0, \beta_{32}^0)$ conformable to the dimensions of $(\mathbf{x}_{11i}, \mathbf{x}_{12i})$. Similarly let $\beta_2 = (\beta_{21}, \beta_{22})$, $\beta_3 = (\beta_{31}, \beta_{32})$.

In this case, with a properly modified rank condition of $\mathbf{z}_{(2),i}$ and $\mathbf{w}_{(2),i}$ which excludes the variables associated with $\mathbf{x}_{12,i}$ and $\sum_{j=1, \neq i}^N g_{ij}\mathbf{x}_{12,j}$, we can identify the coefficients $\beta_{(2)}^0 := (\beta_1^0, \beta_{21}^0, \beta_{31}^0)$ using the same argument that leads to the identification in equation (16). However, we cannot identify the coefficients that correspond to the variable $\mathbf{x}_{12,i}$ and $\sum_{j=1, \neq i}^N g_{ij}\mathbf{x}_{12,j}$. The reason is that controlling the network endogeneity with the control variable (\mathbf{x}_{2i}, a_i) wipes out the information in $(\mathbf{x}_{12,i}, \sum_{j=1, \neq i}^N g_{ij}\mathbf{x}_{12,j})$:

$$\mathbf{x}_{12,i} - \mathbb{E}[\mathbf{x}_{12,i}|\mathbf{x}_{2i}, a_i] = 0 \\ \sum_{j=1, \neq i}^N g_{ij}\mathbf{x}_{12,j} - \mathbb{E}\left[\sum_{j=1, \neq i}^N g_{ij}\mathbf{x}_{12,j} | \mathbf{x}_{2i}, a_i\right] \rightarrow_p 0,$$

where the second convergence holds because $\sum_{j=1, \neq i}^N g_{ij}\mathbf{x}_{12,j}$ converges to a function that depends only on (\mathbf{x}_{2i}, a_i) (see supplementary appendix S.2.3).

Throughout the rest of the paper, when we consider (\mathbf{x}_{2i}, a_i) as a control function, we will without loss of generality apply the restriction in assumption 3 that \mathbf{x}_{1i} and \mathbf{x}_{2i} do not overlap.

V. Estimation

In this section we present two estimation methods. In sections VA and VB, we discuss estimation using a_i and (\mathbf{x}_{2i}, a_i) as control functions, respectively.

A. With a_i as Control Function

The identification scheme of theorem 1 identifies the parameter of interest β^0 with a two-step procedure: (a) control a_i in the outcome equation and yield $y_i - \mathbb{E}(y_i|a_i) = (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i|a_i))'\beta^0 + v_i - \mathbb{E}(v_i)$, and then (b) use $\mathbf{z}_i - \mathbb{E}(\mathbf{z}_i|a_i)$ as IVs for $\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i|a_i)$. If we observe a_i and know the conditional mean functions $\mathbf{h}(a_i) = (h^y(a_i), \mathbf{h}^w(a_i), \mathbf{h}^z(a_i)) := (\mathbb{E}[y_i|a_i], \mathbb{E}[\mathbf{w}_i|a_i], \mathbb{E}[\mathbf{z}_i|a_i])$, then β^0 can be estimated using 2SLS:

$$\hat{\beta}_{2SLS}^{\text{inf}} = \left[\sum_{i=1}^N (\mathbf{w}_i - \mathbf{h}^w(a_i))(\mathbf{z}_i - \mathbf{h}^z(a_i))' \left(\sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}^z(a_i))(\mathbf{z}_i - \mathbf{h}^z(a_i))' \right)^{-1} \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}^z(a_i))(\mathbf{w}_i - \mathbf{h}^w(a_i))' \right]^{-1} \\ \times \left[\sum_{i=1}^N (\mathbf{w}_i - \mathbf{h}^w(a_i))(\mathbf{z}_i - \mathbf{h}^z(a_i))' \left(\sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}^z(a_i))(\mathbf{z}_i - \mathbf{h}^z(a_i))' \right)^{-1} \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}^z(a_i))(y_i - h^y(a_i))' \right]. \quad (17)$$

However, since the individual heterogeneity a_i is not observed and the conditional mean functions $\mathbf{h}(a_i) = (\mathbb{E}(y_i|a_i), \mathbb{E}(\mathbf{w}_i|a_i), \mathbb{E}(\mathbf{z}_i|a_i))$ are not known either, the estimator $\hat{\beta}_{2SLS}^{\text{inf}}$ is not feasible.

A natural implementation of the infeasible estimator $\hat{\beta}_{2SLS}^{\text{inf}}$ is to replace the conditional mean function $\mathbf{h}(a_i)$ with its estimate. Suppose that \hat{a}_i is an estimator of a_i and $\hat{\mathbf{h}}(\hat{a}_i)$ is a nonparametric estimator of $\mathbf{h}(a_i)$. Then we can implement the infeasible estimator $\hat{\beta}_{2SLS}^{\text{inf}}$ with

$$\hat{\beta}_{2SLS} := \left[\sum_{i=1}^N (\mathbf{w}_i - \hat{\mathbf{h}}^w(\hat{a}_i))(\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))' \left(\sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))(\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))' \right)^{-1} \sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))(\mathbf{w}_i - \hat{\mathbf{h}}^w(\hat{a}_i))' \right]^{-1} \\ \sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))(y_i - \hat{h}^y(\hat{a}_i))'. \quad (18)$$

$$\times \left[\sum_{i=1}^N (\mathbf{w}_i - \hat{\mathbf{h}}^w(\hat{a}_i))(\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))' \right. \\ \left. \left(\sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))(\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))' \right)^{-1} \right. \\ \left. \sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}^z(\hat{a}_i))(y_i - \hat{\mathbf{h}}^y(\hat{a}_i))' \right]. \quad (19)$$

See supplementary appendix S.1.1 for more details on the estimator $\hat{\beta}_{2SLS}$.

Estimation of $\mathbf{h}(\cdot)$. We can estimate $\mathbf{h}(\cdot)$ using various standard nonparametric methods. In this paper, we consider a (linear) sieve estimation method.⁷ Suppose that $h^l(a)$ is the l th element in $\mathbf{h}(a)$ for $l = 1, \dots, L$, where L is the dimension of $(y_i, \mathbf{w}_i', \mathbf{z}_i')$. The sieve estimation method assumes that each function $h^l(a)$, $l = 1, \dots, L$ is well approximated by a linear combination of base functions $(q_1(a), \dots, q_{K_N}(a))$:

$$h^l(a) \cong \sum_{k=1}^{K_N} q_k(a) \alpha_k^l, \quad (20)$$

as the truncation parameter $K_N \rightarrow \infty$. A linear sieve (or series) estimator of a function, for example, $\hat{h}^y(\hat{a}_i)$, is the OLS projection of y_i on the sieve basis $\mathbf{q}^K(\cdot) = (q_1(\cdot), \dots, q_{K_N}(\cdot))'$ with \hat{a}_i plugged in,

$$\hat{h}^y(\hat{a}_i) := \mathbf{q}^K(\hat{a}_i)' \left(\sum_{i=1}^N \mathbf{q}^K(\hat{a}_i) \mathbf{q}^K(\hat{a}_i)' \right)^{-1} \sum_{i=1}^N \mathbf{q}^K(\hat{a}_i) y_i.$$

For the regularity conditions of the sieve basis $\mathbf{q}^K(a_i)$, we impose standard conditions such as those proposed by Newey (1997) and Li and Racine (2007). These assumptions ensure that $\sum_{i=1}^N \mathbf{q}^K(a_i) \mathbf{q}^K(a_i)'$ is asymptotically nonsingular and control the rate of approximation of the sieve estimator. These assumptions are formally stated in assumptions 7 and 9 in the appendix.

Additionally, we require that the sieve basis satisfies a Lipschitz condition, which allows us to control for the error introduced by the estimation of a_i with \hat{a}_i in the estimation of $\hat{\beta}_{2SLS}$ ⁸ (see assumptions 8 and 10). As an example, define the polynomial sieve as follows. Let $Pol(K_N)$ denote the space of polynomials on $[-1, 1]$ of degree K_N ,

$$Pol(K_N) = \left\{ v_0 + \sum_{k=1}^{K_N} v_k a^k, a \in [-1, 1], v_k \in \mathbb{R} \right\}.$$

⁷In principle we can use other nonparametric estimation methods such as kernel smoothing or local polynomial methods.

⁸This issue is similar to the two-step series estimation problem in Newey (2009). Other papers that investigated the problem of nonparametric or semiparametric analysis with generated regressors include Ahn and Powell (1993), Mammen, Rothe, and Schienle (2012), Hahn and Ridder (2013), and Escanciano, Jacho-Chávez, and Lewbel (2014), for example.

For any k we have

$$|a_1^k - a_2^k| = k|\tilde{a}^k| |a_1 - a_2| \leq M k |a_1 - a_2|,$$

where $\tilde{a} \in [-1, 1]$ and M is a finite constant.

In sieve estimations, an important issue is choosing the truncation parameter K_N . Well-known procedures for selecting K_N are Mallows's C_p , generalized cross-validation and leave-one-out cross-validation. For more on these methods, see chapter 15.2 in Li and Racine (2007), Wahba (1985), Li (1987), and Hansen (2014). However, these methods are mainly applicable when the observations are cross-sectionally independent, which is not true in our case, especially when the network is dense, as we assume. Developing a data-driven choice of K_N is beyond the scope of this paper, and we leave it for future work.

Estimation of a_i . A desired estimator of a_i should satisfy the following high-level condition:

Assumption 5 (*Estimation of a_i*). We assume that we can estimate a_i with \hat{a}_i such that $\max_i |\hat{a}_i - a_i| = O_p(\zeta_a(N)^{-1})$, where $\zeta_a(N) \rightarrow \infty$ as $N \rightarrow \infty$, satisfying assumption 8 in the appendix.

Here $\zeta_a(N)$ is the order of magnitude that measures the Lipschitz smoothness of the sieve basis. The assumption puts restrictions on the uniform bound of the convergence rate of \hat{a}_i , and we need a more accurate estimator of a_i when the average curvature of the sieve basis is larger.

For the purpose of our paper, any estimation method that yields an estimator \hat{a}_i satisfying the restriction in assumption 5 can be adopted. For example, assuming the parametric specification as in equation (10),

$$d_{ij} = \mathbb{I}(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})' \lambda + a_i + a_j \geq u_{ij}) \quad (21)$$

with regularity conditions of assumption 6 in the appendix, including the error u_{ij} following a logistic distribution, Graham (2017) showed that the joint maximum likelihood estimator that solves

$$(\hat{a}_1, \dots, \hat{a}_N) \\ := \underset{\lambda, (a_1, \dots, a_N)}{\operatorname{argmax}} \left(\sum_{i=1}^N \sum_{j < i} d_{ij} \exp(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})' \lambda + a_i + a_j) \right. \\ \left. - \ln [1 + \exp(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})' \lambda + a_i + a_j)] \right)$$

satisfies

$$\sup_{1 \leq i \leq N} |\hat{a}_i - a_i| \leq O\left(\sqrt{\frac{\ln N}{N}}\right) \quad (22)$$

with probability $1 - O(N^{-2})$. In this case, we have $\zeta_a(N) = \sqrt{\frac{N}{\ln N}}$. Notice that the requirement that the network formation in equation (21) be dense is necessary for \widehat{a}_i to satisfy the desired uniform convergence rate in equation (22). Examples of other estimation methods include Fernández-Val & Weidner (2013), Jochmans (2016, 2018), and Dzemski (2019).

B. With (\mathbf{x}_{2i}, a_i) as Control Function

As we assume in section IVC, we consider the case where \mathbf{x}_{1i} and \mathbf{x}_{2i} do not overlap. When a_i is observed and the conditional expectations $\mathbf{h}_*(\mathbf{x}_{2i}, a_i) = (h_*^y(\mathbf{x}_{2i}, a_i), \mathbf{h}_*^w(\mathbf{x}_{2i}, a_i), \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i)) := (\mathbb{E}(y_i|\mathbf{x}_{2i}, a_i), \mathbb{E}(\mathbf{w}_i|\mathbf{x}_{2i}, a_i), \mathbb{E}(\mathbf{z}_i|\mathbf{x}_{2i}, a_i))$ are known, we can estimate β^0 by the 2SLS similar to $\widehat{\beta}_{2SLS}^{\text{inf}}$ in equation (17):

$$\begin{aligned} & \widehat{\beta}_{2SLS}^{\text{inf}} \\ &= \left[\sum_{i=1}^N (\mathbf{w}_i - \mathbf{h}_*^w(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))' \right. \\ & \quad \left(\sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))' \right)^{-1} \\ & \quad \left. \times \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))(\mathbf{w}_i - \mathbf{h}_*^w(\mathbf{x}_{2i}, a_i))' \right]^{-1} \\ & \times \left[\sum_{i=1}^N (\mathbf{w}_i - \mathbf{h}_*^w(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))' \right. \\ & \quad \left(\sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))' \right)^{-1} \\ & \quad \left. \times \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}_*^z(\mathbf{x}_{2i}, a_i))(y_i - h_*^y(\mathbf{x}_{2i}, a_i))' \right]^{-1}. \quad (23) \end{aligned}$$

When a_i is unknown and \mathbf{x}_{2i} is also used in the control function, under the monotonicity condition of the link formation as in equation (11), we can implement the infeasible estimator using the average node degree without estimating a_i . To be more specific, first we denote

$$\mathbb{P}(d_{ij} = 1|\mathbf{x}_{2i}, a_i) =: \text{deg}(\mathbf{x}_{2i}, a_i) =: \text{deg}_i.$$

Under the monotonicity condition in equation (11), (\mathbf{x}_{2i}, a_i) and $(\mathbf{x}_{2i}, \text{deg}_i)$ are one-to-one. This implies that for any $b_i \in \{y_i, \mathbf{w}_i, \mathbf{z}_i\}$,

$$h_*^b(\mathbf{x}_{2i}, a_i) = \mathbb{E}(b_i|\mathbf{x}_{2i}, a_i) = \mathbb{E}(b_i|\mathbf{x}_{2i}, \text{deg}_i) =: h_*^b(\mathbf{x}_{2i}, \text{deg}_i).$$

Notice that the natural estimator of deg_i is the node degree of i , the number of connections with node (individual) i in the network scaled by the network size:

$$\widehat{\text{deg}}_i := \frac{1}{N-1} \sum_{j=1, j \neq i}^N d_{ij}.$$

Recall that the link d_{ij} is formed by

$$d_{ij} = \mathbb{I}(g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) - u_{ij} \geq 0).$$

Also recall that the unobserved link-specific error terms u_{ij} are assumed to be independent of all the other variables and randomly drawn. Let $\Phi(\cdot)$ be the cdf of u_{ij} . Also let $\pi(\mathbf{x}_2, a)$ be the joint density function of (\mathbf{x}_{2i}, a_i) . Then, for each (\mathbf{x}_{2i}, a_i) , by the WLLN conditioning on (\mathbf{x}_{2i}, a_i) , we have

$$\begin{aligned} \widehat{\text{deg}}_i &:= \frac{1}{N-1} \sum_{j=1, j \neq i}^N \mathbb{I}(g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) - u_{ij} \geq 0) \\ &\rightarrow_p \int \Phi(g(t(\mathbf{x}_{2i}, \mathbf{x}_2), a_i, a)) \pi(\mathbf{x}_2, a) d\mathbf{x}_2 da \\ &= \mathbb{P}(d_{ij} = 1|\mathbf{x}_{2i}, a_i) \\ &=: \text{deg}_i > 0 \end{aligned} \quad (24)$$

as the network size N grows to infinity. Here the limit of the average network $\text{deg}_i > 0$ follows since we assume the network is dense.

This shows that $\widehat{\text{deg}}_i$ can be used as an estimator of deg_i . In fact, we can show that under the regularity conditions in assumption 11 in the appendix, $\sup_i \mathbb{E}[(\sqrt{N}(\widehat{\text{deg}}_i - \text{deg}_i))^{2B}] < \infty$ for any finite integer $B \geq 2$, from which we can deduce that

$$\max_{1 \leq i \leq N} |\widehat{\text{deg}}_i - \text{deg}_i| = O_p(\zeta_{deg}(N)^{-1}), \quad (25)$$

where

$$\zeta_{deg}(N) := o(1)N^{\frac{B-1}{2B}}.$$

This corresponds to the regularity condition in assumption 5.

Suppose that $\mathbf{r}^K(\mathbf{x}_{2i}, \text{deg}_i) = (r_1(\mathbf{x}_{2i}, \text{deg}_i), \dots, r_K(\mathbf{x}_{2i}, \text{deg}_i))'$ is a sieve basis of the unknown function $\mathbf{h}_*(\mathbf{x}_{2i}, a_i)$. For each $b_i \in \{y_i, \mathbf{w}_i, \mathbf{z}_i\}$, a sieve estimator of $h_*^b(\mathbf{x}_{2i}, \text{deg}_i) = \mathbb{E}(b_i|\mathbf{x}_{2i}, a_i)$ is the OLS projection of b_i on $\mathbf{r}^K(\mathbf{x}_{2i}, \widehat{\text{deg}}_i)$ —for example,

$$\begin{aligned} & \widehat{h}_*^y(\mathbf{x}_{2i}, a_i) \\ &= \widehat{h}_{**}^y(\mathbf{x}_{2i}, \text{deg}_i) \\ &= \mathbf{r}^K(\mathbf{x}_{2i}, \widehat{\text{deg}}_i)' \left(\sum_{i=1}^N \mathbf{r}^K(\mathbf{x}_{2i}, \widehat{\text{deg}}_i) \mathbf{r}^K(\mathbf{x}_{2i}, \widehat{\text{deg}}_i)' \right)^{-1} \\ & \quad \sum_{i=1}^N \mathbf{r}^K(\mathbf{x}_{2i}, \widehat{\text{deg}}_i) y_i. \end{aligned}$$

Then we have

$$\begin{aligned} \hat{\beta}_{2SLS} &= \left[\sum_{i=1}^N (\mathbf{w}_i - \hat{\mathbf{h}}_*^w(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))' \right. \\ &\quad \left(\sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))' \right)^{-1} \\ &\quad \times \sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))(\mathbf{w}_i - \hat{\mathbf{h}}_*^w(\mathbf{x}_{2i}, a_i))' \left. \right]^{-1} \\ &\times \left[\sum_{i=1}^N (\mathbf{w}_i - \hat{\mathbf{h}}_*^w(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))' \right. \\ &\quad \left(\sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))' \right)^{-1} \\ &\quad \times \sum_{i=1}^N (\mathbf{z}_i - \hat{\mathbf{h}}_*^z(\mathbf{x}_{2i}, a_i))(y_i - \hat{h}_*^y(\mathbf{x}_{2i}, a_i))' \left. \right]^{-1}. \quad (26) \end{aligned}$$

For more details see supplementary appendix S.1.2.

The two different estimators $\hat{\beta}_{2SLS}$ and $\tilde{\beta}_{2SLS}$ are implemented using different control functions, and these two approaches have their own pros and cons. For $\hat{\beta}_{2SLS}$, a good estimator of a_i is required, which imposes restrictions on the network formation model (9) in the form of equation (10). Compared to this, the estimator $\tilde{\beta}_{2SLS}$ that uses $(\mathbf{x}_{2i}, \text{deg}_i)$ as control functions does not require a restriction like equation (10). It requires only the monotonicity of the net surplus function as in equation (11) of section IIIB. However, $\tilde{\beta}_{2SLS}$ has disadvantages: because it uses x_{2i} as a part of the control function, as discussed in section IVC, this approach cannot identify and estimate the coefficients of the regressor \mathbf{x}_{2i} if \mathbf{x}_{2i} is a relevant regressor of the outcome. Later in section VII, where we present the Monte Carlo simulations, we compare the finite sample properties of $\hat{\beta}_{2SLS}$ and $\tilde{\beta}_{2SLS}$ in both dense and sparse network setups.

VI. Limit Distribution and Standard Error

In this section we present the asymptotic distributions of the two 2SLS estimators, $\hat{\beta}_{2SLS}$ and $\tilde{\beta}_{2SLS}$, and show how to estimate standard errors. We also discuss key technical issues in deriving the limits. All details of the technical derivations and proofs can be found in the appendix.

A. Limiting Distribution and Standard Error of $\hat{\beta}_{2SLS}$

Recall the definitions $h^v(a_i) := \mathbb{E}[y_i | a_i]$, $h^v(a_i) := \mathbb{E}[v_i | a_i]$, $\mathbf{h}^w(a_i) := \mathbb{E}(\mathbf{w}_i | a_i)$, $\mathbf{h}^z(a_i) := \mathbb{E}(\mathbf{z}_i | a_i)$. Define $\eta_i^y := y_i - h^v(a_i)$, $\eta_i^v := v_i - h^v(a_i)$, $\eta_i^w := \mathbf{w}_i - \mathbf{h}^w(a_i)$, $\eta_i^z := \mathbf{z}_i - \mathbf{h}^z(a_i)$.

Let $\boldsymbol{\eta}_N^v = (\eta_1^v, \dots, \eta_N^v)'$ and $\mathbf{H}_N^v(\mathbf{a}_N) = (h^v(a_1), \dots, h^v(a_N))'$. Let $\hat{h}^v(a_i)$, $\hat{\mathbf{h}}^w(a_i)$, and $\hat{\mathbf{h}}^z(a_i)$ denote the sieve estimators of $h^v(a_i)$, $h^w(a_i)$ and $h^z(a_i)$, respectively.

In the appendix, we derive the asymptotic distribution of $\hat{\beta}_{2SLS}$ in three steps. First, we show that the sampling error caused by the use of \hat{a}_i instead of a_i is asymptotically negligible (see lemma 2 in supplementary appendix S.2.1). Next, we control the error introduced by the nonparametric estimation of $h^l(a_i)$, where $l \in \{v, w, z\}$. In lemma 7 in supplementary appendix S.2.2, we show that under the regularity conditions, the estimation error in $\hat{h}^l(a_i)$ vanishes at a suitable rate. Combining these two, we deduce

$$\sqrt{N}(\hat{\beta}_{2SLS} - \hat{\beta}_{2SLS}^{\text{inf}}) = o_p(1).$$

The last step is to derive the limiting distribution of the infeasible estimator $\sqrt{N}(\hat{\beta}_{2SLS}^{\text{inf}} - \beta^0)$. In supplementary appendix S.2.3, we show the following:

$$\frac{1}{N} \sum_{i=1}^N (\mathbf{w}_i - \mathbf{h}^w(a_i))(\mathbf{z}_i - \mathbf{h}^z(a_i))' \xrightarrow{p} \mathbf{S}^{wz}, \quad (27)$$

$$\frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}^z(a_i))(\mathbf{z}_i - \mathbf{h}^z(a_i))' \xrightarrow{p} \mathbf{S}^{zz}, \quad (28)$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}^z(a_i))\eta_i^v \Rightarrow \mathcal{N}(0, \mathbf{S}^{zz\sigma}), \quad (29)$$

where the closed forms of the limits \mathbf{S}^{wz} and \mathbf{S}^{zz} are found in lemma 11 and $\mathbf{S}^{zz\sigma}$ in lemma 12 in the supplementary appendix.

Notice that the derivation of the limiting distribution in equation (29) allows $\eta_i^v = v_i - \mathbb{E}(v_i | a_i)$ to be conditionally heteroskedastic, and so $\sigma^2(\mathbf{x}_i, a_i) := \mathbb{E}[(v_i - \mathbb{E}[v_i | a_i])^2 | \mathbf{x}_i, a_i]$ is allowed to depend on (\mathbf{x}_i, a_i) .

Combining all the limit results leads to the following theorem.

Theorem 3 (Limiting Distribution). Suppose that assumptions 1, 2, 5, 7, 8, and 11(i)–(v) in the appendix hold. Then we have

$$\sqrt{N}(\hat{\beta}_{2SLS} - \beta^0) \Rightarrow \mathcal{N}(0, \Omega),$$

where

$$\begin{aligned} \Omega &= (\mathbf{S}^{wz}(\mathbf{S}^{zz})^{-1}(\mathbf{S}^{wz})')^{-1}(\mathbf{S}^{wz}(\mathbf{S}^{zz})^{-1}\mathbf{S}^{zz\sigma}(\mathbf{S}^{zz})^{-1}(\mathbf{S}^{wz})') \\ &\quad \times (\mathbf{S}^{wz}(\mathbf{S}^{zz})^{-1}(\mathbf{S}^{wz})')^{-1}. \end{aligned} \quad (30)$$

The theorem requires several regularity conditions, which are presented in appendix A.1. In addition to conditions of random sampling of (y_i, \mathbf{x}_i, a_i) in assumption 1 and the full rank condition in assumption 2, we assume conditions that ensure a_i can be consistently estimated and that the error

between $\mathbf{h}(a_i)$ and $\widehat{\mathbf{h}}(\widehat{a}_i)$ converges to 0 at a suitable rate (assumptions 5, 7, and 8). We also impose restrictions on the outcome model, equation (4), and the network formation model, equation (9) (assumption 11). We assume $|\beta_1^0|$ is bounded below 1 so that the spillover effect has a unique solution, and $\|\beta_2^0\|$ is bounded above 0 so that the IVs are strong. We also assume the observables (y_i, \mathbf{x}_i) and \mathbf{t}_{ij} are bounded, and a_i has a compact support in $[-1, 1]$. This boundedness condition is required as a technical regularity condition that simplifies the proofs of the limits in equations (27), (28), and (29), which involves some uniformity in the limit.

The asymptotic variance can be consistently estimated by

$$\widehat{\Omega} = (\widehat{\mathbf{S}}^{\mathbf{wz}}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})')^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}\widehat{\mathbf{S}}^{\mathbf{zz}\sigma}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})') \times (\widehat{\mathbf{S}}^{\mathbf{wz}}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})')^{-1}, \quad (31)$$

where

$$\begin{aligned} \widehat{\mathbf{S}}^{\mathbf{wz}} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{w}_i - \widehat{\mathbf{h}}^{\mathbf{w}}(\widehat{a}_i)) (\mathbf{z}_i - \widehat{\mathbf{h}}^{\mathbf{z}}(\widehat{a}_i))' \\ \widehat{\mathbf{S}}^{\mathbf{zz}} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i - \widehat{\mathbf{h}}^{\mathbf{z}}(\widehat{a}_i)) (\mathbf{z}_i - \widehat{\mathbf{h}}^{\mathbf{z}}(\widehat{a}_i))' \\ \widehat{\mathbf{S}}^{\mathbf{zz}\sigma^2} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i - \widehat{\mathbf{h}}^{\mathbf{z}}(\widehat{a}_i)) (\mathbf{z}_i - \widehat{\mathbf{h}}^{\mathbf{z}}(\widehat{a}_i))' (\widehat{\eta}_i^v)^2, \end{aligned}$$

$$\text{and } \widehat{\eta}_i^v = y_i - \widehat{h}^v(\widehat{a}_i) - (\mathbf{w}_i - \widehat{\mathbf{h}}^{\mathbf{w}}(\widehat{a}_i))' \widehat{\beta}_{2SLs}.$$

B. Limiting Distribution and Standard Error of $\widehat{\beta}_{2SLs}$

The process is analogous to the one presented in the previous section. Again, let b_i^l be the l th element in $(y_i, \mathbf{w}_i', \mathbf{z}_i')'$. Recall the definition that

$$h_*^l(\mathbf{x}_{2i}, a_i) = \mathbb{E}[b_i^l | \mathbf{x}_{2i}, a_i] = \mathbb{E}[b_i^l | \mathbf{x}_{2i}, \text{deg}_i] =: h_{**}^l(\mathbf{x}_{2i}, \text{deg}_i).$$

Further, let $\eta_{*i}^l = b_i^l - h_*^l(\mathbf{x}_{2i}, a_i) = b_i^l - h_{**}^l(\mathbf{x}_{2i}, \text{deg}_i)$, and let $\widehat{h}_{**}^l(\mathbf{x}_{2i}, \text{deg}_i)$ denote a sieve estimator of $h_{**}^l(\mathbf{x}_{2i}, \text{deg}_i)$.

As in the previous section, we derive the asymptotic distribution of $\widehat{\beta}_{2SLs}$ in three steps. First, we show that the error that stems from the use of the estimate $\widehat{\text{deg}}_i$ for deg_i , $\widehat{h}_{**}^l(\mathbf{x}_{2i}, \widehat{\text{deg}}_i) - \widehat{h}_{**}^l(\mathbf{x}_{2i}, \text{deg}_i)$, is asymptotically negligible. In the second step, we control the error introduced by the nonparametric estimation of $h_{**}^l(\mathbf{x}_{2i}, \text{deg}_i)$, $\widehat{h}_{**}^l(\mathbf{x}_{2i}, \text{deg}_i) - h_{**}^l(\mathbf{x}_{2i}, \text{deg}_i)$. This implies

$$\sqrt{N}(\widehat{\beta}_{2SLs} - \widehat{\beta}_{2SLs}^{\text{inf}}) = o_p(1).$$

The last step is to derive the limiting distribution of the infeasible estimator $\sqrt{N}(\widehat{\beta}_{2SLs}^{\text{inf}} - \beta^0)$ by showing

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (\mathbf{w}_i - \mathbf{h}_*^{\mathbf{w}}(\mathbf{x}_{2i}, a_i)) (\mathbf{z}_i - \mathbf{h}_*^{\mathbf{z}}(\mathbf{x}_{2i}, a_i))' &\xrightarrow{p} \bar{\mathbf{S}}^{\mathbf{wz}} \\ \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}_*^{\mathbf{z}}(\mathbf{x}_{2i}, a_i)) (\mathbf{z}_i - \mathbf{h}_*^{\mathbf{z}}(\mathbf{x}_{2i}, a_i))' &\xrightarrow{p} \bar{\mathbf{S}}^{\mathbf{zz}} \\ \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{z}_i - \mathbf{h}_*^{\mathbf{z}}(\mathbf{x}_{2i}, a_i)) \eta_{*i}^v &\Rightarrow \mathcal{N}(0, \bar{\mathbf{S}}^{\mathbf{zz}\sigma}). \end{aligned}$$

Combining all the limit results we have the following theorem.

Theorem 4 (Limiting Distribution). Suppose that assumptions 1, 3, 4, 9, 10, and 11 hold. Then we have

$$\sqrt{N}(\widehat{\beta}_{2SLs} - \beta^0) \Rightarrow \mathcal{N}(0, \bar{\Omega}),$$

where

$$\begin{aligned} \bar{\Omega} &= (\bar{\mathbf{S}}^{\mathbf{wz}}(\bar{\mathbf{S}}^{\mathbf{zz}})^{-1}(\bar{\mathbf{S}}^{\mathbf{wz}})')^{-1}(\bar{\mathbf{S}}^{\mathbf{wz}}(\bar{\mathbf{S}}^{\mathbf{zz}})^{-1}\bar{\mathbf{S}}^{\mathbf{zz}\sigma}(\bar{\mathbf{S}}^{\mathbf{zz}})^{-1}(\bar{\mathbf{S}}^{\mathbf{wz}})') \\ &\quad \times (\bar{\mathbf{S}}^{\mathbf{wz}}(\bar{\mathbf{S}}^{\mathbf{zz}})^{-1}(\bar{\mathbf{S}}^{\mathbf{wz}})')^{-1}. \end{aligned}$$

The asymptotic result in theorem 4 requires the following regularity conditions, which are formally presented in the appendix. First, assumption 3 assumes that the regressors in the outcome equation \mathbf{x}_{1i} and the observables in the network formation \mathbf{x}_{2i} do not overlap. Assumption 4 is a full rank condition for $\bar{\beta}_{2SLs}$. Assumptions 9 and 10 regard the sieve used in constructing the estimator $\widehat{\beta}_{2SLs}$. Compared with the assumptions assumed in theorem 3, theorem 4 does not require the high-level condition of assumption 5 because we do not use an estimator of a_i . Instead it requires an additional restriction that the net surplus function in the link formation be strictly monotonic in a_i conditional on $(\mathbf{x}_{2i}, \mathbf{x}_{2j}, a_j)$, which implies the required monotonicity condition in equation (11).

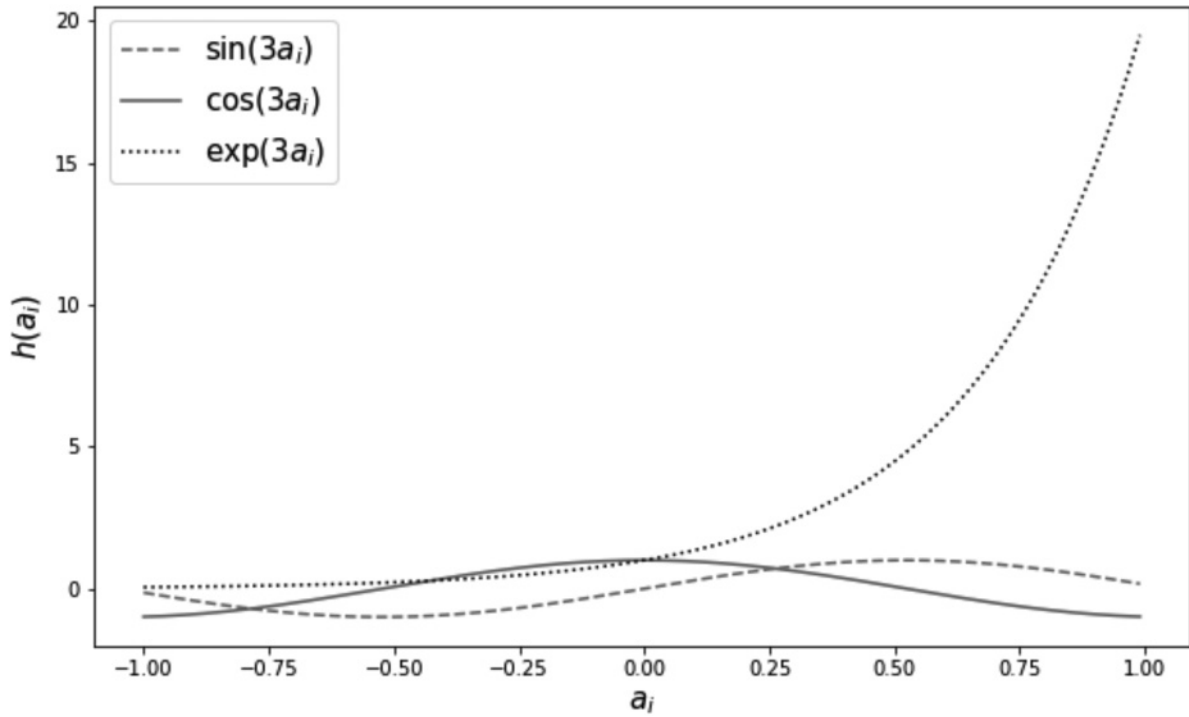
As in the case of $\widehat{\beta}_{2SLs}$, we allow $\eta_{*i}^v = v_i - \mathbb{E}(v_i | \mathbf{x}_{2i}, a_i)$ to be conditionally heteroskedastic, and $\sigma_*^2(\mathbf{x}_i, a_i) := \mathbb{E}[(v_i - \mathbb{E}(v_i | \mathbf{x}_{2i}, a_i))^2 | \mathbf{x}_i, a_i]$ is allowed to depend on (\mathbf{x}_i, a_i) .

The asymptotic variance can be consistently estimated by

$$\begin{aligned} \widehat{\Omega} &= (\widehat{\mathbf{S}}^{\mathbf{wz}}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})')^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}\widehat{\mathbf{S}}^{\mathbf{zz}\sigma}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})') \\ &\quad \times (\widehat{\mathbf{S}}^{\mathbf{wz}}(\widehat{\mathbf{S}}^{\mathbf{zz}})^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})')^{-1}, \end{aligned} \quad (32)$$

where

$$\begin{aligned} \widehat{\mathbf{S}}^{\mathbf{wz}} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{w}_i - \widehat{\mathbf{h}}_{**}^{\mathbf{w}}(\mathbf{x}_{2i}, \widehat{\text{deg}}_i)) (\mathbf{z}_i - \widehat{\mathbf{h}}_{**}^{\mathbf{z}}(\mathbf{x}_{2i}, \widehat{\text{deg}}_i))' \\ \widehat{\mathbf{S}}^{\mathbf{zz}} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i - \widehat{\mathbf{h}}_{**}^{\mathbf{z}}(\mathbf{x}_{2i}, \widehat{\text{deg}}_i)) (\mathbf{z}_i - \widehat{\mathbf{h}}_{**}^{\mathbf{z}}(\mathbf{x}_{2i}, \widehat{\text{deg}}_i))' \end{aligned}$$

FIGURE 1.— $h(a_i)$ FOR SELECTED FUNCTIONAL FORMS OF $h(a_i)$ 

$$\widehat{\mathbf{S}}^{\text{ZZO}^2} = \frac{1}{N} \sum_{i=1}^N (\mathbf{z}_i - \widehat{\mathbf{h}}_{**}^z(\mathbf{x}_{2i}, \widehat{\text{deg}}_i)) (\mathbf{z}_i - \widehat{\mathbf{h}}_{**}^z(\mathbf{x}_{2i}, \widehat{\text{deg}}_i))' \times (\widehat{\eta}_{**i}^v)^2,$$

$$\text{and } \widehat{\eta}_{**i}^v = y_i - \widehat{h}_{**}^y(\mathbf{x}_{2i}, \widehat{\text{deg}}_i) - (\mathbf{w}_i - \widehat{\mathbf{h}}_{**}^w(\mathbf{x}_{2i}, \widehat{\text{deg}}_i))' \widehat{\beta}_{2\text{SLS}}.$$

VII. Monte Carlo

We consider both dense and sparse network Monte Carlo designs. In the dense network, case links are formed according to⁹

$$d_{ij} = \mathbb{I} \{x_{2i}x_{2j}\lambda_d + a_i + a_j - u_{ij} \geq 0\},$$

where $x_{2i} \in \{-1, 1\}$, $\lambda_d = 1$ and u_{ij} follows a logistic distribution. This link rule implies that agents have a strong taste for homophilic matching since $x_{2i}x_{2j}\lambda_d = 1$ when $x_{2i} = x_{2j}$ and $x_{2i}x_{2j}\lambda_d = -1$ when $x_{2i} \neq x_{2j}$.

In the sparse network case, links are formed according to

$$d_{ij} = \mathbb{I} \{(|x_{2i} - x_{2j}| + 3)\lambda_s + a_i + a_j - u_{ij} \geq 0\},$$

with $\lambda_s = -1$. This rule also implies homophily on observable characteristics. Individual-level degree heterogeneity is generated according to

$$a_i = \varphi(\alpha_L \mathbb{I} \{x_{2i} = -1\} + \alpha_H \mathbb{I} \{x_{2i} = 1\} + \xi_i),$$

⁹This follows the approach of Graham (2017).

with $\alpha_L \leq \alpha_H$ and ξ_i a centered beta random variable $\xi_i | x_{2i} \sim \{Beta(\mu_0, \mu_1) - \frac{\mu_0}{\mu_0 + \mu_1}\}$ so that $a_i \in [\alpha_L - \frac{\mu_0}{\mu_0 + \mu_1}, \alpha_H + \frac{\mu_1}{\mu_0 + \mu_1}]$. We choose values of the network formation parameters so that $a_i \in [-1, 1]$. In the main text, we present results based on the following parameter values. In the dense network case, we set $\mu_0 = 1/4$, $\mu_1 = 3/4$, $\alpha_L = \alpha_H = -3/4$, which yields an average node degree = 23 when $N = 100$. The sparse network formation design is generated by setting $\mu_0 = 1$, $\mu_1 = 1$, $\alpha_L = \alpha_H = -1/4$, which gives an average degree = 1.78 when $N = 100$.¹⁰

Individual outcomes are generated according to

$$y_i = \beta_1 \sum_{\substack{j=1 \\ j \neq i}}^N g_{ij} y_j + \beta_2 x_{1i} + \beta_3 \sum_{\substack{j=1 \\ j \neq i}}^N g_{ij} x_{1j} + h(a_i) + \varepsilon_i.$$

In the simulations, we set $\beta_1 = 0.8$, $\beta_2 = \beta_3 = 5$, $x_{1i} = 3q_1 + \cos(q_2)/0.8 + \varepsilon_i$, where $q_1, q_2 \sim \mathcal{N}(x_{2i}, 1)$, and $\varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, 1)$. For $h(a_i)$ we use the following functional forms: $h(a_i) = \exp(3a_i)$, $h(a_i) = \cos(3a_i)$, $h(a_i) = \sin(3a_i)$. A plot of $h(a_i)$ for these functional forms is presented in figure 1. We can see that the exponential function yields a strongly increasing impact on the individual outcome, and with the cosine functions, the returns are increasing up to a certain point and then decreasing; however, the sine function gives a more irregular pattern.

¹⁰Results for fourteen other network formation designs can be found in section S.4 of the online appendix. Most results are similar to the ones presented in the main text.

TABLE 1.—DESIGN 4 DENSE NETWORK: PARAMETER VALUES ACROSS 1000 MONTE CARLO REPLICATIONS WITH $K_N = 4$ AND HERMITE POLYNOMIAL SIEVE

N	100						250						
CF	(0)	(1)	(2)	(3)	(4)	(5)	(0)	(1)	(2)	(3)	(4)	(5)	
$h(a_i) = \exp(a_i)$													
β_1	0.002 (0.010)	0.004 (0.013)	−0.000 (0.015)	−0.000 (0.015)	0.000 (0.024)	−0.000 (0.010)	0.004 (0.009)	0.007 (0.013)	−0.001 (0.015)	−0.001 (0.015)	−0.001 (0.025)	−0.000 (0.009)	(a)
	0.133	0.115	0.056	0.061	0.058	0.058	0.306	0.225	0.057	0.057	0.064	0.050	(b)
β_2	−0.003 (0.031)	−0.004 (0.032)	−0.000 (0.034)	−0.000 (0.033)	0.000 (0.035)	−0.000 (0.031)	−0.002 (0.020)	−0.004 (0.021)	0.000 (0.020)	0.000 (0.020)	−0.000 (0.021)	0.000 (0.020)	(a)
	0.058	0.069	0.074	0.068	0.074	0.057	0.069	0.079	0.055	0.059	0.058	0.061	(b)
β_3	−0.032 (0.178)	−0.048 (0.217)	0.006 (0.251)	0.008 (0.250)	0.006 (0.269)	0.006 (0.174)	−0.066 (0.163)	−0.107 (0.219)	0.009 (0.249)	0.013 (0.248)	0.012 (0.270)	0.009 (0.152)	(a)
	0.078	0.078	0.055	0.060	0.061	0.061	0.156	0.172	0.051	0.054	0.062	0.050	(b)
													(c)
$h(a_i) = \sin(a_i)$													
β_1	−0.008 (0.014)	−0.005 (0.014)	−0.000 (0.016)	−0.000 (0.015)	−0.000 (0.025)	−0.000 (0.011)	−0.015 (0.017)	−0.010 (0.015)	−0.001 (0.015)	−0.001 (0.015)	−0.001 (0.026)	−0.001 (0.010)	(a)
	0.464	0.160	0.058	0.061	0.059	0.045	0.753	0.293	0.054	0.057	0.071	0.053	(b)
β_2	0.007 (0.033)	0.005 (0.034)	−0.001 (0.035)	−0.000 (0.033)	0.000 (0.036)	−0.000 (0.031)	0.007 (0.022)	0.005 (0.022)	−0.000 (0.021)	0.000 (0.020)	−0.000 (0.021)	0.000 (0.020)	(a)
	0.075	0.072	0.067	0.068	0.072	0.060	0.076	0.071	0.056	0.059	0.060	0.055	(b)
β_3	0.113 (0.222)	0.078 (0.231)	0.009 (0.258)	0.008 (0.250)	0.009 (0.277)	0.005 (0.191)	0.236 (0.268)	0.165 (0.249)	0.010 (0.255)	0.013 (0.248)	0.012 (0.276)	0.012 (0.177)	(a)
	0.237	0.100	0.057	0.060	0.053	0.053	0.646	0.248	0.056	0.054	0.055	0.048	(b)
													(c)
$h(a_i) = \cos(a_i)$													
β_1	−0.009 (0.016)	0.004 (0.014)	−0.000 (0.017)	−0.000 (0.015)	0.000 (0.025)	−0.000 (0.010)	−0.017 (0.018)	0.010 (0.016)	−0.000 (0.015)	−0.001 (0.015)	−0.001 (0.026)	−0.001 (0.009)	(a)
	0.459	0.104	0.055	0.061	0.057	0.053	0.745	0.318	0.059	0.057	0.059	0.046	(b)
β_2	0.009 (0.040)	−0.004 (0.034)	−0.000 (0.036)	−0.000 (0.033)	0.001 (0.037)	−0.000 (0.031)	0.008 (0.026)	−0.005 (0.022)	0.000 (0.021)	0.000 (0.020)	0.000 (0.021)	0.000 (0.020)	(a)
	0.075	0.061	0.062	0.068	0.070	0.060	0.084	0.077	0.053	0.059	0.055	0.062	(b)
β_3	0.123 (0.257)	−0.051 (0.232)	0.004 (0.266)	0.008 (0.250)	0.004 (0.286)	0.004 (0.176)	0.264 (0.292)	−0.161 (0.258)	0.008 (0.256)	0.013 (0.248)	0.010 (0.276)	0.011 (0.157)	(a)
	0.224	0.074	0.053	0.059	0.055	0.056	0.640	0.256	0.055	0.054	0.057	0.047	(b)
													(c)

(a) Mean bias, (b) SD, (c) Size, $\beta_1 = 0.8$, $\beta_2 = \beta_3 = 5$. Size is the empirical size of t -test against the truth. CF: control function. (0) None, (1) $\lambda_a \hat{a}_i$, (2) $\hat{h}(\hat{a}_i)$, (3) $\hat{h}(a_i)$, (4) $\hat{h}(\widehat{\deg}_i, x_{2i})$, (5) $h(a_i)$. The network design parameters are $\mu_0 = 0.25$, $\mu_1 = 0.75$, $\alpha_L = -0.75$, $\alpha_H = -0.75$. Average number of links for $N = 100$ is 23.0; for $N = 250$ it is 57.8. Average skewness for $N = 100$ is 0.66; for $N = 250$ it is 0.89. $N = 100$, $\text{corr}(a_i, x_{2i}) = 0.004$, $N = 250$, $\text{corr}(a_i, x_{2i}) = 0.001$. The bias of \hat{a}_i is calculated as $a_i - \hat{a}_i$. For $N = 100$, \hat{a}_i mean bias = 0.018, median bias = 0.008, SD = 0.271. For $N = 250$, \hat{a}_i mean bias = 0.007, median bias = 0.004, SD = 0.167.

We estimate the outcome equation coefficients ($\beta_1, \beta_2, \beta_3$) using the standard 2SLS estimator for peer effects and the Hermite polynomial sieve as well as a polynomial sieve. For the dense network case, we estimate a_i using \hat{a}_i and implement the following control functions: using a control function linear in \hat{a}_i , $\hat{h}(\hat{a}_i)$, $\hat{h}(a_i)$, $\hat{h}(\widehat{\deg}_i, x_{2i})$,¹¹ and $h(a_i)$. For the sparse network case, the estimator of a_i is not reliable,¹² and we implement the following control functions: linear in a_i , $\hat{h}(a_i)$, $\hat{h}(\widehat{\deg}_i, x_{2i})$, and $h(a_i)$. In both the dense and sparse setup, we also implement a benchmark model with no control for the endogeneity of the network.

In the paper, due to space limitations, we present Monte Carlo results obtained using the Hermite polynomial sieve with $K_N = 4$. Specifically, tables 1 and 2 include results for

the dense and sparse network specifications, respectively. Results for the other orders of K_N are not notably different; in the online supplement we provide results for fourteen other network formation designs, for $K_N = 4, 8$ and for the Hermite polynomial and polynomial sieve functions.

We also perform conventional leave-one-out cross validation to find data-dependent K_N (chosen as the K_N that minimizes the RMSE of the prediction based on the leave-one-out estimator (see Li, 1987; Hansen, 2014). We report the statistics on the cross-validation in table 3. The differences in RMSE are very small between the different values of K_N .

Analyzing the Monte Carlo results for the dense network specification in table 1, we can see that, as expected from our asymptotic theories, the control functions $\hat{h}(\hat{a}_i)$ and $\hat{h}(\widehat{\deg}_i, x_{2i})$ perform better than the estimator with a linear control function, as well as the estimator that does not control for the endogeneity of the network in terms of mean bias. This difference is more pronounced in the case when $h(a_i)$ is the sine or cosine function. Both the control for degree approach and the control function that uses $\hat{h}(\hat{a}_i)$ yield a low bias and have the correct size on all coefficients in all cases.

¹¹Note that since x_{2i} is discrete with a finite support, $\{x_1, \dots, x_M\}$, we have $r(x_{2i}, \deg_i) = \sum_{m=1}^M r(x_m, \deg_i) \mathbb{I}\{x_{2i} = x_m\}$. We can then approximate $r(x_{2i}, \deg_i) \simeq \sum_{k=1}^{K_N} \left\{ \sum_{m=1}^M \alpha_{m,k} q_k^d(\deg_i) \mathbb{I}\{x_{2i} = x_m\} \right\}$.

¹²To estimate a_i , we use the JMLE proposed in Graham (2017). As Graham (2017) states, in sparse designs, the JMLE rarely even exists, rendering it unusable in practice when the network is too sparse. See Graham (2017) for more details.

TABLE 2.—DESIGN 4 SPARSE NETWORK: PARAMETER VALUES ACROSS 1,000 MONTE CARLO REPLICATIONS WITH $K_N = 4$ AND HERMITE POLYNOMIAL SIEVE

N	100					250					
	CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)
$\beta_1 = 0.8$		0.001	0.001	0.000	0.000	0.000	0.002	0.003	0.000	−0.000	0.000 (a)
		(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)	(0.004)	(0.002)	(0.003)	(0.002) (b)
		0.089	0.090	0.052	0.056	0.049	0.269	0.257	0.072	0.055	0.064 (c)
$\beta_2 = 5$		−0.001	−0.002	−0.003	−0.002	−0.003	−0.007	−0.008	0.000	0.001	0.001 (a)
		(0.039)	(0.039)	(0.033)	(0.041)	(0.032)	(0.027)	(0.027)	(0.021)	(0.025)	(0.021) (b)
		0.043	0.046	0.065	0.061	0.060	0.078	0.084	0.055	0.066	0.049 (c)
$\beta_3 = 5$		−0.004	−0.004	−0.002	0.002	−0.002	−0.027	−0.028	−0.001	−0.000	−0.001 (a)
		(0.076)	(0.077)	(0.066)	(0.075)	(0.065)	(0.063)	(0.064)	(0.052)	(0.058)	(0.051) (b)
		0.034	0.038	0.063	0.063	0.047	0.085	0.090	0.056	0.068	0.060 (c)
$h(a_i) = \sin(a_i)$											
$\beta_1 = 0.8$		−0.000	0.000	0.000	0.000	0.000	−0.002	−0.000	0.000	−0.000	0.000 (a)
		(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002) (b)
		0.059	0.048	0.052	0.057	0.051	0.170	0.068	0.072	0.059	0.071 (c)
$\beta_2 = 5$		−0.007	−0.002	−0.003	−0.002	−0.003	0.005	0.001	0.000	0.001	0.000 (a)
		(0.039)	(0.032)	(0.033)	(0.041)	(0.032)	(0.026)	(0.022)	(0.021)	(0.025)	(0.021) (b)
		0.052	0.061	0.066	0.062	0.059	0.083	0.061	0.055	0.073	0.048 (c)
$\beta_3 = 5$		−0.001	−0.001	−0.002	0.002	−0.002	0.016	−0.001	−0.001	−0.000	−0.001 (a)
		(0.078)	(0.067)	(0.066)	(0.076)	(0.065)	(0.064)	(0.052)	(0.052)	(0.058)	(0.051) (b)
		0.059	0.053	0.063	0.067	0.049	0.079	0.057	0.056	0.065	0.057 (c)
$h(a_i) = \cos(a_i)$											
$\beta_1 = 0.8$		0.001	0.001	0.000	0.000	0.000	0.002	0.002	0.000	−0.000	0.000 (a)
		(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002) (b)
		0.073	0.081	0.052	0.053	0.049	0.197	0.216	0.072	0.067	0.068 (c)
$\beta_2 = 5$		−0.002	−0.002	−0.003	−0.002	−0.003	−0.005	−0.006	0.000	0.001	0.000 (a)
		(0.038)	(0.038)	(0.033)	(0.041)	(0.032)	(0.025)	(0.025)	(0.021)	(0.025)	(0.021) (b)
		0.047	0.051	0.066	0.061	0.062	0.062	0.074	0.055	0.065	0.047 (c)
$\beta_3 = 5$		−0.003	−0.003	−0.002	0.002	−0.002	−0.020	−0.022	−0.001	0.000	−0.001 (a)
		(0.073)	(0.073)	(0.066)	(0.074)	(0.065)	(0.061)	(0.062)	(0.052)	(0.059)	(0.051) (b)
		0.038	0.036	0.063	0.065	0.049	0.069	0.079	0.056	0.070	0.062 (c)

(a) Mean bias, (b) SD, (c) Size, $\beta_1 = 0.8$, $\beta_2 = \beta_3 = 5$. Size is the empirical size of t -test against the truth. CF: control function. (0) None, (1) $\lambda_a a_i$, (2) $\hat{h}(a_i)$, (3) $\hat{h}(\widehat{deg}_i, x_{2i})$, (4) $h(a_i)$. The network design parameters are $\mu_0 = 1.00$, $\mu_1 = 1.00$, $\alpha_L = -0.25$, $\alpha_H = -0.25$. Average number of links for $N = 100$ is 1.8; for $N = 250$ it is 4.5. Average skewness for $N = 100$ is 0.81; for $N = 250$ it is 0.62. Size is the empirical size of t -test against the truth. $N = 100$, $corr(a_i, x_{2i}) = -0.001$, $N = 250$, $corr(a_i, x_{2i}) = -0.002$.

In the simulations we also implemented the control function $\widehat{h}(a_i)$, that is, using the true a_i instead of \widehat{a}_i . These results are very similar to the ones obtained using $\widehat{h}(\widehat{a}_i)$, which is in line with the estimator \widehat{a}_i having a very low bias, as detailed in the table footnotes. This suggests that the approach of using $\widehat{h}(\widehat{a}_i)$ as a control function works very well when a highly precise estimator of a_i is available (e.g., when the network size N is large.).

Looking at table 2 and the results for the sparse design, we can see that the control for degree approach performs very well across all functional forms of $h(a_i)$. In the sparse setup, the bias of all estimates, including those that do not control for the endogeneity of the network, is small. However, the size of the no control and linear control estimates is not correct. If a precise estimator of a_i is available, the control function $\widehat{h}(a_i)$ also performs well with low bias and correct size in all cases.

Table 3 shows that the performance of the estimators does not differ notably for different values of K_N . As for the choice of K_N we present in the tables, we have run simulations for a range of values of K_N , and the results did not differ significantly. As deriving a theory for a data-driven choice of K_N is beyond the scope of this paper, for applied researchers,

we suggest estimating the model over a range of K_N and seeing whether the results vary significantly. As shown in our Monte Carlo simulations, the control function approach yields results robust to the choice of K_N for different nonlinear functions.

VIII. Conclusion

In this paper, we show that whenever the network is likely endogenous, it is important to control for this endogeneity when estimating peer effects. Failing to control for the endogeneity of the connections matrix in general leads to biased estimates of peer effects. We show that under specific assumptions, we can use the control function approach to deal with the endogeneity problem. We assume that unobserved individual characteristics directly affect link formation and individual outcomes. We leave the functional form through which unobserved individual characteristics enter the outcome equation unspecified and estimate it using a nonparametric approach. The estimators we propose are easy to use in applied work, and Monte Carlo results show that they perform well compared to a linear control function estimator. Erroneously assuming that unobserved characteristics enter

TABLE 3.—CROSS-VALIDATION RESULTS: PARAMETER VALUES ACROSS 1,000 MONTE CARLO REPLICATIONS FOR DENSE NETWORK DESIGN 4 AND HERMITE POLYNOMIAL SIEVE

N		100						250					
K_N		3	4	5	6	7	8	3	4	5	6	7	8
$\beta_0 - \hat{\beta}_0$		Control function: $\hat{h}(a_i)$											
$\exp(a_i)$	mean	1.287	1.247	1.279	1.280	1.288	1.264	1.172	1.181	1.188	1.200	1.185	1.181
	median	0.576	0.551	0.561	0.562	0.569	0.568	0.530	0.534	0.534	0.538	0.531	0.531
	SD	1.864	1.813	1.887	1.905	1.889	1.816	1.673	1.691	1.702	1.733	1.698	1.681
	iqr	1.553	1.499	1.532	1.543	1.546	1.537	1.427	1.436	1.442	1.450	1.442	1.441
$\cos(a_i)$	mean	1.877	1.898	1.883	1.866	1.925	1.884	1.793	1.810	1.809	1.797	1.800	1.795
	median	0.921	0.931	0.916	0.922	0.940	0.916	0.896	0.904	0.901	0.897	0.896	0.889
	SD	2.528	2.538	2.528	2.490	2.608	2.537	2.351	2.380	2.380	2.362	2.373	2.373
	iqr	2.333	2.402	2.357	2.344	2.407	2.357	2.274	2.282	2.292	2.274	2.274	2.261
$\sin(a_i)$	mean	1.433	1.450	1.454	1.452	1.490	1.483	1.360	1.375	1.362	1.369	1.367	1.375
	median	0.647	0.653	0.652	0.665	0.675	0.666	0.624	0.632	0.620	0.631	0.619	0.625
	SD	2.050	2.071	2.072	2.051	2.144	2.140	1.911	1.936	1.915	1.920	1.946	1.940
	iqr	1.730	1.762	1.783	1.773	1.803	1.799	1.666	1.680	1.672	1.680	1.673	1.673
$\beta_0 - \hat{\beta}_0$		Control function: $\hat{h}(\widehat{\deg}_i, a_i)$											
$\exp(a_i)$	mean	1.930	1.908	1.879	1.898	1.977	1.891	1.625	1.584	1.666	1.705	1.636	1.601
	median	0.784	0.775	0.749	0.756	0.797	0.762	0.700	0.682	0.714	0.711	0.701	0.689
	SD	3.124	3.105	3.203	3.198	3.371	3.015	2.437	2.409	2.538	2.716	2.482	2.444
	iqr	2.181	2.166	2.085	2.120	2.193	2.152	1.926	1.860	1.954	1.965	1.922	1.889
$\cos(a_i)$	mean	2.555	2.522	2.527	2.535	2.576	2.570	2.225	2.220	2.268	2.244	2.219	2.236
	median	1.137	1.135	1.125	1.148	1.154	1.142	1.060	1.054	1.062	1.056	1.043	1.043
	SD	3.854	3.931	3.956	3.893	3.900	4.009	3.088	3.106	3.235	3.159	3.143	3.189
	iqr	3.039	2.957	2.956	2.990	3.066	3.014	2.745	2.724	2.763	2.749	2.713	2.721
$\sin(a_i)$	mean	2.058	2.033	2.053	1.996	2.093	2.085	1.755	1.799	1.768	1.742	1.805	1.845
	median	0.861	0.838	0.860	0.846	0.877	0.878	0.780	0.797	0.773	0.774	0.782	0.795
	SD	3.244	3.392	3.216	3.119	3.315	3.317	2.560	2.677	2.622	2.574	2.769	2.935
	iqr	2.380	2.317	2.383	2.327	2.416	2.416	2.108	2.144	2.105	2.080	2.137	2.156

The statistics are based on conventional leave one out cross-validation.

the outcome equation in a linear fashion can lead to a serious bias in the estimated parameters.

REFERENCES

- Abadie, Alberto, and Guido W. Imbens, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica* 74 (2006), 235–267.
- Ahn, Hyuntaik, and James L. Powell, "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics* 58:1–2 (1993), 3–29.
- Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone, "Parametric and Semiparametric IV Estimation of Network Models with Selectivity," Einaudi Institute for Economics and Finance technical report (2015).
- Auerbach, Eric, "Identification and Estimation of Models with Endogenous Network Formation," working paper (2016).
- Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson, "The Diffusion of Microfinance," *Science* 341:6144 (2013), 1236498.
- Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides, "Identification of Social Interactions" (pp. 853–964), in Jess Benhabib, Alberto Bisin, and Matthew Jackson, eds., *Handbook of Social Economics* (Amsterdam: Elsevier, 2011).
- Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman, "Linear Social Interactions Models," *Journal of Political Economy* 123 (2015), 444–496.
- Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of Peer Effects through Social Networks," *Journal of Econometrics* 15:1 (2009), 41–55.
- Brock, William A., and Steven N. Durlauf, "Interactions-Based Models" (pp. 3297–3380), in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics* (Amsterdam: Elsevier, 2001).
- Chen, Mingli, Iván Fernández-Val, and Martin Weidner, *Nonlinear Factor Models for Network and Panel Data* (2014), arXiv:1412.5647, 2014.
- De Paula, Aureo, "Econometrics of Network Models" (pp. 268–323), in David M. Kreps and Kenneth F. Wallis, eds., *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress* (Cambridge: Cambridge University Press, 2017).
- De Weerd, Joachim, and Marcel Fafchamps, "Social Identity and the Formation of Health Insurance Networks," *Journal of Development Studies* 47 (2011), 1152–1177.
- Ductor, Lorenzo, Marcel Fafchamps, Sanjeev Goyal, and Marco J. van der Leij, "Social Networks and Research Output," this REVIEW 96 (2014), 936–948.
- Dzemska, Andreas, "An Empirical Model of Dyadic Link Formation in a Network with Unobserved Heterogeneity," this REVIEW 101 (2019), 763–776.
- Epplé, Dennis, and Richard E. Romano, "Peer Effects in Education: A Survey of the Theory and Evidence" (pp. 1053–1163), in Jess Benhabib, Alberto Bisin, and Matthew Jackson, eds., *Handbook of Social Economics* (Amsterdam: Elsevier, 2011).
- Escanciano, Juan, Carlos David T. Jacho-Chávez, and Arthur Lewbel, "Uniform Convergence of Weighted Sums of Non and Semiparametric Residuals for Estimation and Testing," *Journal of Econometrics* 178 (2014), 426–443.
- Fafchamps, Marcel, and Flore Gubert, "Risk Sharing and Network Formation," *American Economic Review* 97 (2007), 75–79.
- Fernández-Val, I., and Martin Weidner, *Individual and Time Effects in Nonlinear Panel Models with Large N, T* (2013), arXiv:1311.7065.
- Goldsmith-Pinkham, Paul, and Guido W. Imbens, "Social Networks and the Identification of Peer Effects," *Journal of Business and Economic Statistics* 31 (2013), 253–264.
- Graham, Bryan S., "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers" (pp. 965–1052), in Jess Benhabib, Alberto Bisin, and Matthew Jackson, eds., *Handbook of Social Economics* (Amsterdam: Elsevier, 2011).
- , "An Econometric Model of Network Formation with Degree Heterogeneity," *Econometrica* 85 (2017), 1033–1063.
- Hahn, Jinyong, and Geert Ridder, "Asymptotic Variance of Semiparametric Estimators with Generated Regressors," *Econometrica* 81 (2013), 315–340.

- Hansen, Bruce E., "Nonparametric Sieve Regression: Least Squares, Averaging Least Squares, and Cross-Validation" (pp. 215–248), in Jeffrey Racine, Liangjun Su, and Aman Ullah, eds., *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* (Oxford: Oxford University Press, 2014).
- Heckman, James J., Hidehiko Ichimura, and Petra Todd, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies* 65 (1998), 261–294.
- Hsieh, Chih-Sheng, and Lung-Fei Lee, "A Social Interactions Model with Endogenous Friendship Formation and Selectivity," *Journal of Applied Econometrics* 31 (2016), 301–319.
- Jackson, Matthew O., "A Survey of Network Formation Models: Stability and Efficiency" (pp. 11–49), in Gabrielle Demange and Myrta Wooders, eds., *Group Formation in Economics: Networks, Clubs, and Coalitions* (Cambridge: Cambridge University Press, 2005).
- Jochmans, Koen, "Modified-Likelihood Estimation of the B-Model," Sciences Po Department of Economics technical report (2016).
- , "Semiparametric Analysis of Network Formation," *Journal of Business and Economic Statistics* 36 (2018), 705–713.
- Johnsson, Ida, and Hyungsik Roger Moon, "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach," University of Southern California working paper (2019), <http://www-bcf.usc.edu/moon/>.
- Kelejian, Harry H., and Ingmar R. Prucha, "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics* 17 (1998), 99–121.
- Lee, Lung-Fei, "Best Spatial Two Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances," *Econometric Reviews* 22 (2003), 307–335.
- , "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics* 140 (2007a), 333–374. doi:10.1016/j.jeconom.2006.07.001.
- , "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models," *Journal of Econometrics* 137 (2007b), 489–514.
- Lee, Lung-fei, Xiaodong Liu, and Xu Lin, "Specification and Estimation of Social Interaction Models with Network Structures," *Econometrics Journal* 13 (2010), 145–176.
- Li, Ker-Chau, "Asymptotic Optimality for C_p , C_l , Cross-Validation and Generalized Cross-Validation: Discrete Index Set," *Annals of Statistics* 15 (1987), 958–975.
- Li, Qi, and Scott Jeffrey Racine, *Nonparametric Econometrics: Theory and Practice* (Princeton, NJ: Princeton University Press, 2007).
- Mammen, Enno, Christoph Rothe, and Melanie Schienle, "Nonparametric Regression with Nonparametrically Generated Covariates," *Annals of Statistics* 40 (2012), 1132–1170.
- Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* 60 (1993), 531–542.
- , "Economic Analysis of Social Interactions," NBER technical report (2000).
- Newey, Whitney K., "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics* 79 (1997), 147–168.
- , "Two-Step Series Estimation of Sample Selection Models," *Econometrics Journal* 12:s1 (2009), S217–S229.
- Powell, James, *Semiparametric Estimation of Bivariate Latent Variable Models* (Madison: University of Wisconsin–Madison, Social Systems Research Institute, 1987).
- Qu, Xi, and Lung-Fei Lee, "Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix," *Journal of Econometrics* 184 (2015), 209–232.
- Robinson, Peter, "Root-N-Consistent Semiparametric Regression," *Econometrica: Journal of the Econometric Society* 56 (1998), 931–954.
- Shalizi, Cosma Rohilla, "Comment on 'Why and When "Flawed" Social Network Analyses Still Yield Valid Tests of No Contagion,'" *Statistics, Politics, and Policy* 3:1 (2012), 5.
- Sheng, Shuyang, "Identification and Estimation of Network Formation Games," Unpublished manuscript (2012).
- Wahba, Grace, "A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem," *Annals of Statistics* 13 (1985), 1378–1402.
- Weinberg, Bruce A., "Social Interactions with Endogenous Associations," NBER working paper 13038 (2007).

Appendix

In this section we introduce the assumptions that are required for the two asymptotic results, theorem 3 for $\hat{\beta}_{2SLS}$ and theorem 4 for $\tilde{\beta}_{2SLS}$. The proof of theorem 3 is available in the supplementary appendix, which is available in Johnsson and Moon (2019). Since the proof of theorem 4 is similar to that of theorem 3, we provide only a sketch of the proof of theorem 4 in the supplementary appendix.

A. Assumptions

In this section we introduce the assumptions used in the proof of theorem 3. First, we introduce a set of sufficient conditions under which we can estimate a_i satisfying the conditions in assumption 5. This assumption corresponds to assumptions 1, 2, 3, and 5 of Graham (2017).

Assumption 6 (Sufficient Conditions for assumption 5). (i) $\mathbf{t}_{ij} = \mathbf{t}_{ji}$. (ii) $u_{ij} \sim i.i.d.$ for all ij a logistic distribution. (iii) The supports of λ , \mathbf{t}_{ij} , a_i are compact.

The next four assumptions are about the sieves used in the semiparametric estimators. The first two are for $\hat{\beta}_{2SLS}$ and the next two for $\tilde{\beta}_{2SLS}$.

Assumption 7 (Sieve). For every K_N there is a nonsingular matrix of constants \mathbf{B} such that for $\tilde{\mathbf{q}}^{K_N}(a) = \mathbf{B}\mathbf{q}^{K_N}(a)$, we assume the following. (i) The smallest eigenvalue of $\mathbb{E}[\tilde{\mathbf{q}}^{K_N}(a_i)\tilde{\mathbf{q}}^{K_N}(a_i)']$ is bounded away from 0 uniformly in K_N . (ii) There exists a sequence of constants $\zeta_0(K_N)$ that satisfy the condition $\sup_{a \in \mathcal{A}} \|\tilde{\mathbf{q}}^{K_N}(a)\| \leq \zeta_0(K_N)$, where K_N satisfies $\zeta_0(K_N)^2 K_N/N \rightarrow 0$ as $N \rightarrow \infty$. (iii) For $f(a)$ being an element of $\mathbf{h}(a) = (E[y_i|a_i = a], E[\mathbf{z}_i|a_i = a], E[\mathbf{w}_i|a_i = a])$, there exists a sequence of $\alpha_{K_N}^f$ and a number $\kappa > 0$ such that

$$\sup_{a \in \mathcal{A}} \|f(a) - \mathbf{q}^{K_N}(a)' \alpha_{K_N}^f\| = O(K_N^{-\kappa})$$

as $K_N \rightarrow \infty$. (iv) As $N \rightarrow \infty$, $K_N \rightarrow \infty$ with $\sqrt{N}K_N^{-\kappa} \rightarrow 0$ and $K_N/N \rightarrow 0$.

Assumption 8 (Lipschitz Condition). The sieve basis satisfies the following condition: there exists a positive number $\zeta_1(k)$ such that

$$\|\mathbf{q}_k(a) - \mathbf{q}_k(a')\| \leq \zeta_1(k)\|a - a'\| \quad \forall k = 1, \dots, K_N,$$

with $\frac{1}{\zeta_a(N)^2} \sum_{k=1}^{K_N} \zeta_1^2(k) = o(1)$ and $\zeta_0(K_N)^6 \left(\frac{1}{\zeta_a(N)^2} \sum_{k=1}^{K_N} \zeta_1^2(k) \right) = o(1)$.

In our paper, we use the following sieves for the Monte Carlo simulations:

(i) Polynomial: For $|a| \leq 1$, define

$$Pol(K_N) = \left\{ v_0 + \sum_{k=1}^{K_N} v_k a^k, a \in [-1, 1] v_k \in \mathbb{R} \right\}.$$

(ii) The Hermite Polynomial sieve: For $|a| \leq 1$, define

$$H\text{Pol}(K_N) = \left\{ \sum_{k=1}^{K_N+1} v_k H_k(a) \exp\left(\frac{-a^2}{2}\right), \right. \\ \left. a \in [-1, 1], v_k \in \mathbb{R} \right\},$$

$$\text{where } H_k(a) = (-1)^k e^{a^2} \frac{d^k}{da^k} e^{-a^2}.$$

For the polynomial sieve, it is known that $\zeta_0 = O(K_N)$ (e.g., Newey, 1997). Then, since $\zeta_1(k) = O(k)$, $\sum_{k=1}^{K_N} \zeta_1^2(k) = O(K_N^3)$. Hence, the conditions that must be satisfied for the polynomial sieve are $K_N^3/N \rightarrow 0$ and $\sqrt{N}K_N^{-\kappa} \rightarrow 0$. Further, when $\zeta_a(N)^2 = \frac{N}{\ln N}$, we need $\zeta_a(N)^{-2} O(K_N^9) = o(1)$.

The next two assumptions are for the sieves used in $\tilde{\beta}_{2SLs}$. These assumptions modify assumptions 7 and 8.

Assumption 9 (Sieve). For every K_N , there is a nonsingular matrix of constants \mathbf{B} such that for $\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, \text{deg}_i) = \mathbf{B}\mathbf{r}^{K_N}(\mathbf{x}_{2i}, \text{deg}_i)$. We assume the following. (i) The smallest eigenvalue of $\mathbb{E}[\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, \text{deg}_i)\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, \text{deg}_i)']$ is bounded away from 0 uniformly in K_N . (ii) There exists a sequence of constants $\zeta_{0**}(K_N)$ that satisfy the condition $\sup_{(\mathbf{x}_{2i}, \text{deg}_i) \in \mathcal{S}} \|\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, \text{deg}_i)\| \leq \zeta_{0**}(K_N)$, where K_N satisfies $\zeta_{0**}(K_N)^2 K_N/N \rightarrow 0$ as $N \rightarrow \infty$, and \mathcal{S} is the domain of $(\mathbf{x}_{2i}, \text{deg}_i)$. (iii) For $f(\mathbf{x}_{2i}, \text{deg}_i)$ being an element of $\mathbf{h}_{**}(\mathbf{x}_{2i}, \text{deg}_i) = (\mathbb{E}[y_i|\mathbf{x}_{2i}, \text{deg}_i], \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, \text{deg}_i], \mathbb{E}[\mathbf{w}_i|\mathbf{x}_{2i}, \text{deg}_i])$, there exists a sequence of $\mathbf{y}_{K_N}^f$ and a number $\kappa > 0$ such that

$$\sup_{(\mathbf{x}_{2i}, \text{deg}_i) \in \mathcal{S}} \|f - \mathbf{r}^{K_N'} \mathbf{y}_{K_N}^f\| = O(K_N^{-\kappa})$$

as $K_N \rightarrow \infty$. (iv) As $N \rightarrow \infty$, $K_N \rightarrow \infty$ with $\sqrt{N}K_N^{-\kappa} \rightarrow 0$ and $K_N/N \rightarrow 0$.

Recall that $\sup_i |\widehat{\text{deg}}_i - \text{deg}_i| = O(\zeta_{\text{deg}}(N)^{-1})$ with $\zeta_{\text{deg}}(N) = o(1)N^{\frac{B-1}{2B}}$ for some integer $B \geq 2$.

Assumption 10 (Lipschitz). For $\zeta_{0**}(K_N)$ being the constant from assumption 10, there exists a positive number $\zeta_{1**}(k)$ such that

$$\|\mathbf{r}_k(\mathbf{x}_{2i}, \text{deg}_i) - \mathbf{r}_k(\mathbf{x}_{2i}, \text{deg}'_i)\| \\ \leq \zeta_{1**}(k) \|\text{deg}_i - \text{deg}'_i\| \quad \forall k = 1, \dots, K_N$$

with $\zeta_{\text{deg}}(N)^{-2} \sum_{k=1}^{K_N} \zeta_{1**}^2(k) = o(1)$ and $\zeta_{0**}(K_N)^6 \left(\zeta_{\text{deg}}(N)^{-2} \sum_{k=1}^{K_N} \zeta_{1**}^2(k) \right) = o(1)$.

The next assumptions restrict the models of the outcome in equation (4) and the network formation of equation (9). We need assumption 11 to derive the limiting distribution of $\tilde{\beta}_{2SLs}$ in theorem 3.

Assumption 11. We assume the following: (i) The true coefficients satisfy $|\beta_1^0| \leq 1 - \epsilon$ and $\|\beta_2^0\| \geq \epsilon$ for some small ϵ . (ii) The parameter set \mathbb{B} for β is bounded. (iii) The observables (y_i, \mathbf{x}_i) are bounded. The unobserved characteristic a_i has a compact support in $[-1, 1]$. (iv) The network formation error u_{ij} has an unbounded full support \mathbb{R} . (v) The net surplus of the network $g(\mathbf{t}_{ij}, a_i, a_j)$ is bounded by a finite constant, where $\mathbf{t}_{ij} := \mathbf{t}(\mathbf{x}_{2i}, \mathbf{x}_{2j})$. (vi) The net surplus of the network $g(\mathbf{t}_{ij}, a_i, a_j)$ is a strictly monotonic function of a_i for fixed $(\mathbf{x}_{2i}, \mathbf{x}_{2j})$ and a_j .

Condition (i) is standard in the linear-in-means peer effect literature. As discussed in the main text, the condition $|\beta_1^0| \leq 1 - \epsilon$ is required for a unique solution of the spillover effect. We need the restriction $\|\beta_2^0\| > \epsilon$ for the IVs to be strong. The boundedness conditions in (ii) and (iii) are important technical assumptions for asymptotics, which require some uniform convergence. Also, these conditions imply key regularity conditions for the CLT. Conditions (vi) and (v) assume that the network is dense and $0 < \kappa \leq \mathbb{E}[d_{ij} = 1] \leq \bar{\kappa} < 1$.

Finally, notice that assumption 11 allows $v_i - \mathbb{E}(v_i|a_i)$ to be conditionally heteroskedastic, and so $\sigma^2(\mathbf{x}_i, a_i) := \mathbb{E}[(v_i - \mathbb{E}[v_i|a_i])^2|\mathbf{x}_i, a_i]$ depends on (\mathbf{x}_i, a_i) . This is also true for $v_i - \mathbb{E}(v_i|a_i)$.