Social Interactions in Multilayered Observational Networks

Juan Estrada*

August 15, 2022

Abstract

This paper proposes a new method to identify and estimate the parameters of an extension of a linear model of peer effects where individuals form different types of social and professional connections that can affect their outcomes. A stylized model provides a theoretical framework for the peer effects linear specification and the main identifying assumptions, which are used to provide identification results in a setting that allows all layers in the multilayer network to be endogenous. I show that identifying heterogeneous network effects is possible under the assumption that the dependence between individuals in the population is characterized by a stochastic process where dependence vanishes in the network space. I offer a novel multilayer measure of distance that provides a source of exogenous variation used to form identifying moment conditions. I propose a Generalized Method of Moments estimator that is consistent and asymptotically normal at the standard rate, for which I characterize the asymptotic variance-covariance matrix that considers the intrinsic network dependence among individuals. A Monte Carlo experiment confirms the desirable finite properties of the proposed estimator, and an empirical application finds positive and significant peer effects in citations from a multilayer network of professional connections.

Keywords: Multilayer Networks, ψ-dependence, Peer Effects, Network Endogeneity, GMM.
JEL Classifications: C13, C31, C51

*Department of Economics, Emory University, Rich Building 306, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA. ✉: jjestra@emory.edu and 🔗: www.juanestrada.info/. This article was part of my dissertation titled "Causal Inference in Multilayered Networks." I want to thank all my committee members for their continuous support and encouragement. Special thanks to Bruno Ferman, Guido Kuersteiner, Alejandro Sanchez Becerra, Pietro Spini, Davide Viviano, and seminar participants at the European Winter Meetings of the Econometric Society, 16th Economics Graduate Students' Conference at Washington University in St. Louis, LACEA LAMES Annual-Meeting 2021, and the Young Academics Networks Conference for helpful comments. I acknowledge the invaluable help from Venkat Balasubramanian from the Bank of Canada, who implemented the algorithms required to conduct the simulation analysis and the empirical application.

1 Introduction

Economic models where social interactions influence individual behavior are becoming increasingly popular in the literature. The so-called Linear-in-Means (LiM) model is the most widely used tool in applied work to estimate the effects of peers' behaviors and characteristics on individual outcomes (see Section 3.1 in de Paula, 2017, pp. 275-289). The challenges to identify the parameters of the LiM model are widely recognized in the econometrics of networks literature. In particular, the outstanding identification issue in the field is how to address the endogenous network formation problem (Jackson et al., 2017). Recent methods designed to solve this issue are built under the standard assumption that only one social or professional network exists and they generally require explicit network formation models, see, e.g., Johnsson and Moon (2019) for estimation of network effects and Graham (2017, 2020) for estimation of network formation models. However, empirical work has shown that different types of connections such as classmates, neighbors, friends, or coauthors can create peer effects (Miguel and Kremer, 2004; Conley and Udry, 2010; Bursztyn et al., 2014; Ductor et al., 2014; Zacchia, 2019; De Giorgi et al., 2020).

This paper proposes a novel method to employ multilayer network data to identify and estimate the parameters of an extension of the LiM model. The proposed approach can handle both the reflection problem and the issue of correlated effects while at the same time allowing for different types of social connections to generate network effects. The standard LiM model assumes full observability of one network whose potential impacts on individuals' outcomes are summarized by the peer and contextual effects. These parameters capture the effects of the outcomes and characteristics of an individual's peers on her own outcome. I propose an extension where I assume full observability of M different networks (also called layers here) that can produce M potentially different peer and contextual effects. I call this model the *Multilayer Linear-in-Means model* (MLiM hereafter). This generalization is meaningful because it nests standard models relevant to applied work and relaxes the assumption of *monolayer* network effects. I provide a game theoretical stylized model in the spirit of Blume et al. (2015) that offers a theoretical foundation for the MLiM model and the main identifying assumptions derived from the multilayer network formation model.

As originally proposed by Kelejian and Prucha (1998), I solve the reflection problem in the MLiM by using the exclusions restrictions generated by a multilayer network data structure that is not fully connected. However, the multiple endogenous layers in the MLiM model make solving the issue of the correlated effect more complex as the network structure cannot be used directly as an instrument. Instead of imposing explicit structural restrictions on the multilayer network formation process, it is assumed that a $\psi-$dependent stochastic process characterizes the dependence of individuals in the population (Doukhan and Louhichi, 1999; Kojevnikov et al., 2020). Based on the $\psi-$dependence characterization, I impose the main identifying assumption, referred to Weak Neighborhood Dependence (WND). This assumption guarantees that individuals' dependence dies out when their distance in the multilayer network space increases. Imposing the $\psi-$dependence assumption on any network that happens with positive probability allows constructing moment conditions to separately identify the potential M different peer and contextual effects after controlling for the presence of correlated effects (endogenous network formation).

This article proposes an innovation to the idea of ψ-dependence to accommodate multilayer network data structures by providing a novel *multilayer network measure of distance* that considers both the standard monolayer geodesic distance and the number of edge-type changes. The idea of a measure of distance that incorporates the complete information provided by the multilayer network data is at the core of my identification, estimation, and inference strategies. It allows taking advantage of local multilayer network structures nearly independent from each other to form moment conditions and incorporate network dependence for inference. In the multilayer network context, edge types refer to the nature of the social or professional relationship connecting two nodes. I interpret the edge-type changes as reducing the dependence between any two nodes faster than standard monolayer paths. Empirical articles have used a similar idea to argue that, for instance, the correlation of individual i's characteristics with her coworker spouse's characteristics is lower than that between her and her coworker's coworker (De Giorgi et al., 2020). Nicoletti et al. (2018) and Nicoletti and Rabe (2019) use similar arguments in the context of a multilayer network composed by friends and neighbors connections.

Current methods to identify the LiM model's parameters in the presence of correlated effects generally require estimating a monolayer network formation model. This paper abstracts away from network formation estimation and is agnostic about the underlying process generating the network. Instead, it uses changes in the characteristics of individuals who are sufficiently far in the multilayer network space as a source of exogenous variation to construct moment conditions for identification and estimation. These characteristics can include local shocks to certain parts of the multilayer network, as in De Giorgi et al. (2020). To the best of my knowledge, this is the first work to formalize the use of multilayer network data structures to identify the network effects parameters of an extension of the linear-in-means model with flexible assumptions on the network formation process. Prior to this paper, Chan et al. (2022) formalized how to use multiplex networks to identify a monolayer LiM model's parameters when the network of interest forms endogenously, and Manta et al. (2022) provided identification results for a MLiM model with exogenous network formation. The results in this paper are more general because it does not require any layer in the multilayer network data to be strictly exogenous. The proposed estimator collapses to the estimators in Chan et al. (2022) and Manta et al. (2022) when additional exogeneity assumptions are imposed in the model.

The potential network endogeneity is modeled by assuming the existence of a joint distribution that allows for correlation between the idiosyncratic errors, the multilayer network, and the model's regressors. Instead of imposing parametric assumptions on the joint distribution, I use the WND assumption on the *dependence coefficients*. Intuitively, any underlying multilayer network formation process will not set individuals with similar characteristics at a long multilayer network distance. In addition to the weak dependence assumption, the relevance condition imposes restrictions on the multilayer network to guarantee enough identifying variation. These conditions are used to show that the parameters of the MLiM model are point identified. An example of a network formation model that contains primitive assumptions that are sufficient for the WND assumption to hold is provided. The proposed network generating process is not the only model that can generate WND, but it serves as an example to motivate the plausibility of the WND high level assumption.

Based on the moment conditions used for identification, I propose a Generalized Method of Moments (GMM) estimation procedure that is consistent and asymptotically normal. The sample consists of n individuals drawn from an arbitrarily large population characterized by a joint distribution of the errors, the multilayer network, and the regressors. Limiting distributions are studied when $n \to \infty$. The linearity assumption of the MLiM model guarantees that the resulting GMM estimator has a closed-form solution. The asymptotic results show that the variance-covariance matrix of the GMM estimator differs from the standard sandwich formulas because it considers the network dependence between individuals and their heterogeneity in the identifying information they provide about the population. A HAC estimator of the variance-covariance matrix is used in the same spirit of Newey and West (1987), Conley (1999) and Kojevnikov et al. (2020). The derived asymptotic variance-covariance matrix in this paper formally explains the anecdotal finding that Monte Carlo variability increases with network density (see, e.g., Bramoullé et al., 2009). Intuitively, higher network density reduces the possibility of forming moment conditions which reduces the identification power in the population and increases the variance of the estimator in the sample. These results are new and relevant for correct inference in empirical work estimating network effects. A Monte Carlo experiment based on an exemplary network formation model confirms the desirable finite properties of the proposed estimator when the assumption that individuals' dependence decreases with their distance in the multilayer network space holds.

To show the importance of taking into account network endogeneity when dealing with observational data, I present an empirical application to publication outcomes in Economics. The use of web scraping and existing data on authors' research fields, education, and employment history allows the creation of four types of professional ties among scholars: co-authorship, alumni, advisors, and colleagues connections. The multilayer network data is used to uncover positive and significant peer effects in citations from the co-authorship network among articles published by these scholars. However, peer effects from any of the other types of networks included in the estimating model are not found. This result is interpreted as emphasizing the importance of a network that guarantees a direct communication channel between authors instead of other professional networks that may generate fewer interpersonal interactions. The empirical application also shows that the OLS estimator can be severely biased when trying to estimate network effects without considering the potential network endogeneity of the layers. Positive results of research teams that are gender diverse on the quality of a paper measured in terms of citation outcomes are also found, after controlling for other articles' characteristics such as number of pages, number of bibliographic references, and various network fixed effects.

Related Literature

This paper provides new insights on current identification results in the literature. The multilayer network data structure is general enough to cover cases such as the non-overlapping network structure used in De Giorgi et al. (2020) and the multiplex structure in Chan et al. (2022). The proposed framework collapses to the reduced form version of Zacchia's (2019) model in the monolayer case when the researcher is willing to assume that the endogenous peer effect coefficient is zero. In that sense, this paper presents a general theory of using multilayer network data to estimate network effects in linear models. The commonality between the articles mentioned above emerges from the idea of keeping the network formation process unspecified. Recent papers augment the standard linear-in-means model to include specific generating mechanisms for the network formation process, see e.g., Goldsmith-Pinkham and Imbens (2013), Qu and Lee (2015) and Johnsson and Moon (2019). In general, these network formation models are difficult to estimate and involve additional assumptions such as the absence of strategic interactions on individuals' utilities of forming peers. For a complete discussion on the importance of strategic interactions to network formation models' point-identification see Graham (2017), De Paula et al. (2018) and Graham and Pelican (2020).

This work relates with the broad reduced form literature on social interactions. It is most closely associated with observational studies which aim to identify both endogenous peer effects and contextual effects. There have been two recent approaches to deal with endogenous network formation in observational studies: quasi randomization and structural endogeneity (Bramoullé et al., 2020). This article separates from the structural endogeneity approach to a significant extent. For instance, the proposed model does not have to assume a particular source of unobserved heterogeneity, and it does not require network formation estimation. The framework is also distinguished from models found in the literature using natural or artificial experiments to randomize peers because it is explicitly assumed that the networks can form endogenously. Thus, the proposed approach is closer to the literature using random shocks (experimental or quasi-experimental) on the regressors for identification. The reason for this is that the use of individual characteristics of distant nodes as a source of exogenous variation can be interpreted as a partial population experiment, see Moffitt (2001) and Kuhn et al. (2011). Also closely related is the approach proposed by Kuersteiner and Prucha (2020), which extends the standard linear-in-means model to include panel data. The two methods relate in that they present an extension of the standard linear model with additional data to provide new identification and estimation results.

The structure of the paper is as follows. Section 2 introduces the multilayer network data structure and its representation in terms of adjacency matrices. Section 3 introduces the MLiM model and provides conditions under which it has a solution in terms of regressors, errors, and layers. The model section also provides some examples where the multilayer network data structure has been used in empirical work. Section 4 presents conditions for the parameters of the MLiM to be uniquely recovered (point identification) from the joint distribution characterizing the infinite population of interest. Appendix A in the Online Appendix outlines a two-stage game including a formation model and a game of social interactions under which the main identifying assumptions are satisfied. Section 5 describes the proposed GMM estimation procedure, its asymptotic distribution, and how to calculate valid asymptotic standard errors. Section 6 presents a Monte Carlo simulation study, while section 7 presents the empirical application to publication outcomes in economics. Finally, Section 8 provides a summary of the main findings and their implications, while section 9 presents the proof for the main results. The Online Appendix contain the theoretical background with the stylized model of social interactions, the proofs for intermediate results, proposed algorithms, and data construction and robustness results in Appendix A, Appendix B and Appendix C, respectively.

2 Preliminaries

This section introduces the background and notation necessary to develop the framework for identification and estimation. Following Boccaletti et al.'s (2014) notation, a multilayer network is a pair $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ where $\mathcal{G} = \{G_m; m \in \{1, \dots, M\}\}$ is a set of graphs $G_m = (V_m, E_m)$. For each graph m, V_m and E_m represent the set of nodes and edges, respectively. In principle, the graphs in $\mathcal{G}$ are allowed to be directed or undirected, weighted or unweighted but are assumed not to have self cycles. The graphs forming the set $\mathcal{G}$ are known as the multilayer network layers. The set of edges E_m is known as *intralayer connections*. To complete the multilayer network structure's characterization, let $\mathcal{C}$ be the set of interconnections between nodes of different layers G_m and G_s with $m \neq s$ known as *crossed layers* and constructed as $\mathcal{C} = \{E_{m,s} \subseteq V_m \times V_s; m, s \in \{1, \dots, M\}, m \neq s\}$. The elements of each set $E_{m,s}$ are known as the *interlayer connections* of $\mathcal{M}$. To accommodate the multilayer network data structure into the MLiM model, I represent the multilayer network by the adjacency matrix of each layer G_m. I denote each adjacency matrix by $\mathbf{W}_m = [w_{m;i,j}]$, where $w_{m;i,j} = \rho_{m;i,j}$ if $(v_{m;i}, v_{m;j}) \in E_m$ and 0 otherwise. The constant $\rho_{m;i,j} \in (0, 1]$ represents the weights on the (i, j)th connection, which may or may not sum up to one. Using this notation, the next section introduces the MLiM model and provides conditions under which the model has a solution in terms of errors, regressors, and layers' adjacency matrices.

3 Multilayer Linear-in-Means (MLiM) Model

The object of study is a MLiM model where agents can create connections in more than one social or professional aspect. The MLiM can be derived as a best response function of a structural game of social interactions with a quadratic utility (Blume et al., 2015). Appendix A in the Online Appendix provides the theoretical framework motivating the MLiM used as a basis to analyze identification and estimation. The model is composed of a collection $\mathcal{I}_N$ of N economic agents (N is allowed to be arbitrarily large), where a set of Q characteristics $\mathbf{x}_{N,i}$ describes each individual. The choices and characteristics of a person's peers can influence her decision-making process. Individuals' social interactions can be embedded into a multilayer network $\mathcal{M}_N$ composed by a set $\mathcal{G}_N$ of M graphs. Based on the Bayesian Game of Social Interactions in Appendix A, it is possible to write the optimal choice (outcome) of an individual as

$$y_{N,i} = \alpha^0 + \sum_{m=1}^{M}\sum_{j\neq i} w_{N,m;i,j} y_{N,j}\beta_m^0 + \sum_{m=1}^{M}\sum_{j\neq i} w_{N,m;i,j}\mathbf{x}_{N,j}^\top \boldsymbol{\delta}_m^0 + \mathbf{x}_{N,i}^\top \boldsymbol{\gamma}^0 + \varepsilon_{N,i}, \tag{1}$$

where $j \in \{1,\ldots,N\}$, $w_{N,m}$ represents the adjacency matrix of layer m, $\varepsilon_{N,i}$ is an unobserved shock which may include unobserved idiosyncratic characteristics relevant to determine the outcome y_i, or believes about others' private types in a setting of incomplete information as in Blume et al. (2015) (see Appendix A). The coefficients $(\beta_{0,m}, \boldsymbol{\delta}_{0,m})$ represent the social effects for network $m \in M$ while $\boldsymbol{\gamma}_0$ captures the direct effects. The zero superscript is used to emphasise that $[\beta_1^0, \ldots \beta_M^0, \boldsymbol{\delta}_1^0, \ldots, \boldsymbol{\delta}_M^0, \alpha^0, \boldsymbol{\gamma}^0]$ is the true parameter vector. The model can be written in matrix form as

$$\mathbf{y}_N = \alpha^0 \boldsymbol{\iota}_N + \left(\sum_{m=1}^{M} \beta_m^0 \mathbf{W}_{N,m}\right)\mathbf{y}_N + \sum_{m=1}^{M} \mathbf{W}_{N,m}\mathbf{X}_N \boldsymbol{\delta}_m^0 + \mathbf{X}_N \boldsymbol{\gamma}^0 + \boldsymbol{\varepsilon}_N. \tag{2}$$

This article focuses on the representation of the multilayer network as the set of layers' adjacency matrices. This approach is motivated by the fact the proposed MLiM model does not consider interlayer connections to affect individuals' outcomes. The multilayer measure of distance define in section 4 can handle data structures where the multilayer network could contain interlayer connections. Then, so long as the model in (2) is correctly specified -in the sense that intralayer edges do not generate network effects- the identification idea proposed in this paper still works. In potential settings where the interlayer connections exist and can generate network effects, it is still possible to use this paper's identification approach. The only required modification is to include the regressors associated with interlayer connections on the right-hand side of the MLiM by using, for instance, an interlayer adjacency matrix.

Let $\mathbf{S}(\boldsymbol{\beta}^0, \mathcal{M}_N) = \mathbf{I}_N - \sum_{m=1}^{M} \beta_m^0 \mathbf{W}_{N,m}$. The model described by equation (2) has a solution for a given $\mathbf{X}_N$, $\boldsymbol{\varepsilon}_N$, and $\mathcal{M}_N$ if the matrix $\mathbf{S}(\boldsymbol{\beta}^0, \mathcal{M}_N)$ has an inverse. Lemma 2 in Appendix B in the Online Appendix shows that the parametric restrictions in Assumption 1 are sufficient to guarantee the existence of $\mathbf{S}^{-1}(\boldsymbol{\beta}^0, \mathcal{M}_N)$. Regarding the adjacency matrices' characteristics, the invertibility result in Lemma 2 only requires the assumption of no cycles for each adjacency matrix m. Thus, it covers cases of directed, undirected, weighted, or unweighted graphs. When the adjacency matrices of the layers are weighted such that $\|\mathbf{W}_{N,m}\|_\infty = 1$ for all m, the condition in Assumption 1 reduces to $|\beta_1^0| + \cdots + |\beta_M^0| < 1$, which is a generalization of a familiar assumption on the peer effects coefficient that is customary in the literature when $m = 1$ (see, e.g., Kelejian and Prucha (1998), Kelejian and Prucha (2001), and Lee (2007a) in spacial econometrics, Lee (2007b) and Bramoullé et al. (2009) in econometrics of networks, to mention some).

Assumption 1 (Invertibility) *The adjacency matrix of layer m has no cycles, i.e., $w_{N,m;i,i} = 0$ for all $m = 1, \ldots, M$ and $i = 1, \ldots, N$. The peer effects coefficients associated with the layers $\mathbf{W}_{N,1}, \mathbf{W}_{N,2}, \ldots, \mathbf{W}_{N,M}$ are such that $|\beta_1^0| \|\mathbf{W}_{N,1}\|_\infty + \cdots + |\beta_M^0| \|\mathbf{W}_{N,M}\|_\infty < 1$, where $\|\mathbf{W}_{N,m}\|_\infty = \sup_{i \in \mathcal{I}_N} \sum_{j=1}^N |w_{N,m;ij}|$ and $m = 1, \ldots M$.*

If Assumption 1 is satisfied, it follows that the solution for equation (2) can be written in terms of $\mathbf{X}_N$, $\boldsymbol{\varepsilon}_N$, and $\mathcal{M}_N$ as follows

$$\mathbf{y}_N = \mathbf{S}^{-1}(\boldsymbol{\beta}^0, \mathcal{M}_N) \left(\alpha^0 \boldsymbol{\iota}_N + \sum_{m=1}^{M} \mathbf{W}_{N,m} \mathbf{X}_N \boldsymbol{\delta}_m^0 + \mathbf{X}_N \boldsymbol{\gamma}^0 + \boldsymbol{\varepsilon}_N \right), \tag{3}$$

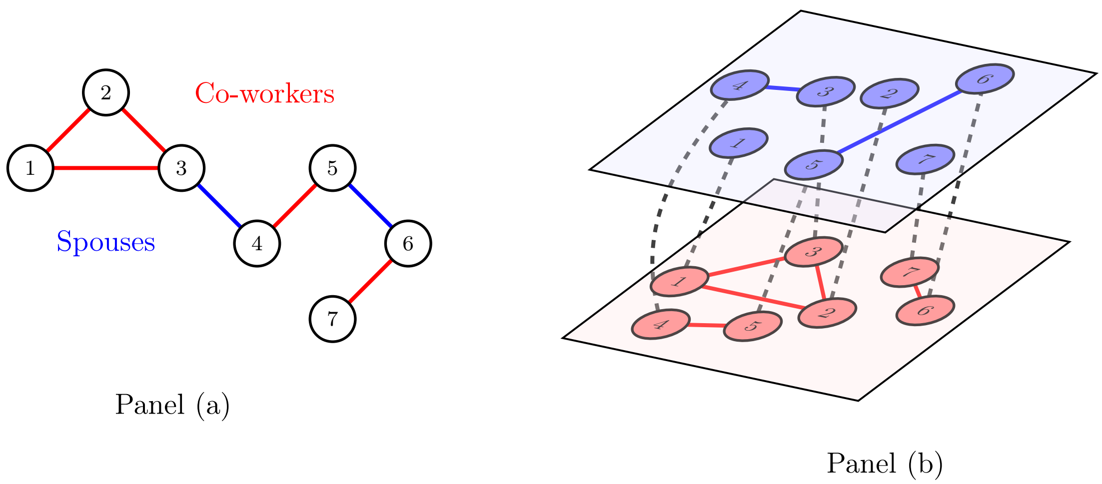
which makes explicit the correlation between $\mathbf{W}_{N,m}\mathbf{y}_N$ and $\boldsymbol{\varepsilon}_N$ for all m. In addition to the endogeneity of the variable $\mathbf{W}_{N,m}\mathbf{y}_N$, this article allows for a general type of dependence between observable characteristics, networks, and unobserved shocks. In practice, endogeneity can arise when individuals form connections in the networks $\{G_{N,m}\}_{m=1}^M$ based on observed and unobserved characteristics correlated with the outcome $\mathbf{y}_N$. If individuals sort themselves into groups following preferences such as observed and unobserved homophily, the network structures in $\{G_{N,m}\}_{m=1}^M$ will be correlated with the errors, and they will also induce correlation between $\mathbf{X}_N$ and $\boldsymbol{\varepsilon}_N$. Recent work investigating the estimation of social effects under network endogeneity has been based on approaches that explicitly model the network formation process, see, e.g., Johnsson and Moon (2019). This article, however, abstract away from explicit network formation assumptions. Instead, as detailed in section 4, the main idea to overcome this endogeneity issue is to form independent multilayer sub-network structures that allow creating moment conditions to identify the parameters in equation (3) separately.

One legitimate question regarding the relevance of the MLiM mode is whether multilayer network data are available in empirical contexts. Before detailing the main identification approach, the next subsection provides a couple of examples that showcase both the availability and versatility of the multilayer data structure.

3.1 Examples

Example 1 (non-overlapping networks): De Giorgi et al. (2020) studied consumption network effects in a context where co-workers are the relevant reference group. Acknowledging the potential network endogeneity, the authors used what they define as a non-overlapping network structure to form valid peer consumption instruments. Essentially, the data structure contains two types of connections: spouses and co-workers. Figure 3 presents a minimal example of their primary data structure. Panel (a) in Figure 1 shows a flat representation of the network data structure where the connections between co-workers and spouses are depicted in blue and red, respectively. Panel (b) shows the multilayer representation of the network in Panel (a). As mentioned before, intralayer edges are the only relevant type of connections. Interlayer edges do not provide relevant information as they only connect a node

Figure 1: De Giorgi et al.'s (2020) Data Structure Represented as a Multilayer Network



Note: Panel (a) displays an example of the non-overlapping network data structure in De Giorgi et al. (2020) which is a modification of the example presented in Figure 2 of their article. The blue edges represent co-workers' connections while the red edges illustrate spouses relationships. Panel (b) shows the representation of the same graph as a multilayer network. The non-overlapping is particular to De Giorgi et al. (2020). However, **my method can also accommodate** links between two people across multiple layers, i.e. if the spouses are also coworkers.

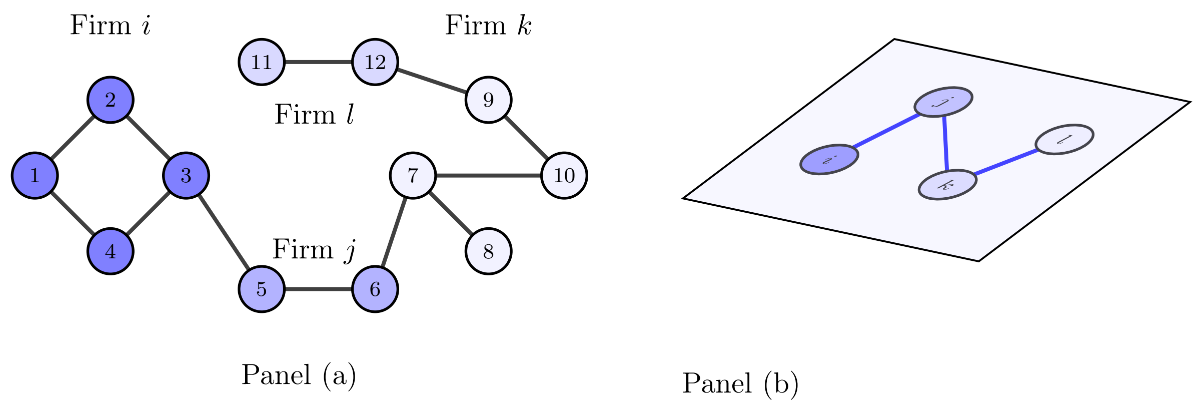
in one layer to itself in the other layer. The non-overlapping is particular to De Giorgi et al. (2020). However, the method proposed in the present paper can also accommodate links between two people across multiple layers, i.e. if the spouses are also coworkers.

Example 2 (monolayer network): Zacchia (2019) studied knowledge spillovers generated by interactions between inventors of different firms. The estimation procedure is based on a monolayer network where two firms are connected if they employ inventors who have collaborated before. Zacchia (2019) is a relevant example because it provides an analytical framework to understand how network endogeneity complicates the identification of contextual effects in a linear model where the endogenous peer effect parameter is not of interest. Figure 2 presents an simplified example of the network data structure in Zacchia (2019). This example explicitly highlights the fact that any monolayer network can be represented as a multilayer network by setting $M = 1$.

4 Identification

The identification strategy in this article abstracts away from a network formation model. Instead, I assume that the matrix of observable characteristics $\mathbf{X}_N$, the multilayer network $\mathcal{M}_N$, and the vector of shocks $\boldsymbol{\varepsilon}_N$ are random draws from a joint distribution $\mathcal{F}(\mathbf{X}_N, \mathcal{M}_N, \boldsymbol{\varepsilon}_N)$. This joint distribution could reflect a potential correlation between $\mathcal{M}_N$ and $\boldsymbol{\varepsilon}_N$ caused by a strategic network formation

Figure 2: Zacchia's (2019) Monolayer Network Data Structure



Note: Panel (a) displays the monolayer data structure in Zacchia (2019). The nodes represent inventors who are connected by an edge if they have worked together in a project before. Nodes of the same color belong to the same firm. Panel (b) represents the monolayer firms network. Two firms i and j are connected if at least one of the inventors working for i is connected to an inventor working for j.

characterized by unobserved homophily. Correlation between $\mathbf{X}_N$ and $\boldsymbol{\varepsilon}_N$ is also allowed as matching based on homophily could induce network dependence across observed and unobserved characteristics. The identification approach is based on imposing the restriction that functionals of network data become uncorrelated as observations becomes distant in the network space characterized by $\mathcal{M}_N$.

In particular, I assume that implicit network formation processes or the potential existence of local network shocks induce correlation patterns that decrease with the individuals' distance in the network space. Thus, there exists a level of distance between individuals such that their observed and unobserved characteristics are not correlated. The second relevant assumption imposes restrictions on the interlayer network dependence. Intuitively, dependence decreases faster if the distance path between two individuals involves changes in the social or professional connections. It is possible to formalize this intuition using the of ψ-dependence framework proposed by Doukhan and Louhichi (1999) and Kojevnikov et al. (2020). The ψ-dependence approach requires the availability of a metric that can characterize the distance between individuals in the multilayer network space. Kojevnikov et al. (2020) uses the *minimum path lengths* (or geodesic distance) which is the standard monolayer measure of distance. However, my proposed identification idea uses multilayer network data, which requires to propose a measure of distance that takes into account the existence of different types of connections.

4.1 Multilayer Measure of Distance

It is possible to extend the geodesic distance definition to the multilayer framework. Following Boccaletti et al. (2014), for a given multilayer $\mathcal{M}_N = (\mathcal{G}_N, \mathcal{C}_N)$, a *multilayer walk* of length $q-1$ can be defined as a sequence of edges $\{v_{1;m_1}, v_{2;m_2}, \ldots, v_{q;m_q}\}$ where $m_1, m_2, \ldots, m_q \in \{1, \ldots, M\}$ and two adjacent nodes in the sequence are connected by an edge belonging to the set $\{E_{N,1}, \ldots, E_{N,M}\} \bigcup \mathcal{C}_N$. In words, a walk connects nodes $v_{1;m_1}$ and $v_{q;m_q}$ (which are allowed to be in different layers) trough a sequence of nodes that can be connected by either intralayer or interlayer edges. For instance, Alice knows Bob from work (layer 1), and Bob knows Cassey because they are neighbors (layer 2). Alice and Cassey do not work in the same place. However, they are at a distance 2 once we consider all the layers in the multilayer network. A *multilayer path* is then defined as a *multilayer walk* where each node is only visited once, and a *multilayer minimum path lengths* is the shortest *multilayer path* connecting two nodes, and it is denoted by $d_N^*(i, j)$. The star superscript is used to emphasize that from all the possible paths connecting i and j, those at distance $d_N^*(i, j)$ are the shortest.

In addition to shortest path length, the proposed measure of distance also takes into account the number of edge type changes in a path connecting two individuals. To formalize this idea, let $\mathcal{D}_N(i, j; d)$ be the set of all possible paths $p_N(i, j)$ -which can include paths across multiple layers- of length d connecting individuals i and j. Based on the set of all possible paths, the set of all possible shortest paths are defined as $\mathcal{D}_N(i, j) = \arg\min_{d \in \mathbb{N}_+} \mathcal{D}_N(i, j; d)$ (where $\mathbb{N}_+$ is the set of positive natural numbers). For instance, the minimum path length between individuals 1 and 5 in Figure 1 is given by $d_7^*(1, 5) = 3$ and the set $\mathcal{D}_7(1, 5) = \{(1, 3, 4, 5)\}$ is a singleton in this case. For any path $p_N(i, j) \in \mathcal{D}_N(i, j; d)$, define $c_N(i, j; p)$ as path p's number of edge type changes. In the example, the total number of edge type changes associated with the shortest path $(1, 3, 4, 5)$ is given by $c_7(1, 5; (1, 3, 4, 5)) = 2$. In general, the set $\mathcal{D}_N(i, j)$ does not need to be a singleton.

There could be different combinations of nodes connecting two individuals with the same number of edges. To incorporate the intuition that less edge type changes are associated with shorter distances, let $c_N^*(i, j) = \min_{p \in \mathcal{D}_N(i,j)} c_N(i, j; p)$ be the minimum number of edge-type changes of the shortest paths connecting i and j. Let $\mathcal{D}_N^*(i, j)$ be the set of all paths $p_N(i, j)$ of length $d_N^*(i, j)$ and total edge changes $c_N^*(i, j)$. The sets $\mathcal{D}_N(i, j)$ and $\mathcal{D}_N^*(i, j)$ are such that $\mathcal{D}_N^*(i, j) \subseteq \mathcal{D}_N(i, j)$ as $\mathcal{D}_N^*(i, j)$ only consider the shortest paths with the minimum number of edge type changes (the set $\mathcal{D}_N^*(i, j)$ does not have to be a singleton either). I call the paths $p_N(i, j) \in \mathcal{D}_N^*(i, j)$ *multilayer shortest paths.* Having described these objects, I now define the *multilayer measure of distance* as

$$d_N^{\mathcal{M}}(i, j) = d_N^*(i, j) + \tau_{i,j} c_N^*(i, j), \tag{4}$$

where $\tau_{ij} > 1$ is a constant that captures the intuition that the distance between two individuals i and j (and consequently their levels of dependence) increases (reduces) faster when the shortest path connecting them involves edge-type changes. In other words, this measure of distance penalizes edge type changes more than shortest path lengths. I characterize the value of τ_{ij} in Proposition 1 below. This penalizing idea has been used in the physics literature of multilayer networks. Kivela et al. (2014) suggests that it is natural to hypothesize that paths containing only one edge type may have different *lengths* than those containing more than one type. In the context of potential restrictions to the $\mathcal{F}$ distribution, I argue that it is reasonable to think that for any given path of length $d_N^*(i,j)$, the dependence between the observed and unobserved characteristics of individuals i and j should decrease with the number of different edge types connecting them. Intuitively, it may be more likely to find similarities between two indirectly connected coworkers than between a person and her spouse's coworker (De Giorgi et al., 2020). Based on the pairwise measure of distance, it is possible to define the distance between sets of nodes. Following Kojevnikov et al. (2020), for $a, b \in \mathbb{N}_+$ the distance between two sets A and B with a and b nodes respectively is given by

$$d_N(A,B) = \min_{i \in A} \min_{j \in B} d_N^{\mathcal{M}}(i,j). \tag{5}$$

Through the identification and estimation sections, I will be using different sets that contain groups of nodes that are at different intralayer or multilayer distances. The following definition collects all the distance sets that will be used in the following sections.

Definition 1 (Distance Sets) *Consider the distance measures $d_N^{\mathcal{M}}(i,j)$, $d_N(A,B)$ and $d_N^*(i,j,m)$, where $d_N^*(i,j,m)$ is the geodesic distance between individuals i and j only considering connections in layer $m \in \{1,\dots,M\}$. Consider the sets: (i) $\mathcal{P}_N^+(a,b,d) = \{(A,B) : A, B \subset \mathcal{I}_N, |A| = a, |B| = b,$ and $d_N(A,B) \geq d\}$ containing groups of nodes at distance of at least d from each other, (ii) $\mathcal{P}_N^-(a,b,d) = \{(A,B) : A, B \subset \mathcal{I}_N, |A| = a, |B| = b,$ and $d_N(A,B) \leq d\}$ containing groups of nodes at distance of at most d from each other, (iii) $\mathcal{P}_N(a,b,d) = \{(A,B) : A, B \subset \mathcal{I}_N, |A| = a, |B| = b,$ and $d_N(A,B) = d\}$ the set associated with groups of nodes at distance d from each other. The associated set that contain all nodes at a certain distance of node i are $\mathcal{P}_n^+(i,d) = \{j \in \mathcal{I}_n : d_n^{\mathcal{M}}(i,j) \geq d\}$, $\mathcal{P}_n(i,d) = \{j \in \mathcal{I}_n : d_n^{\mathcal{M}}(i,j) = d\}$ and $\mathcal{P}_n^-(i,d) = \{j \in \mathcal{I}_n : d_n^{\mathcal{M}}(i,j) \leq d\}$. The same notation applies for sets based on the interlayer connections of layer m by adding the index: $\mathcal{P}_n^+(i,d,m)$, $\mathcal{P}_n(i,d,m)$ and $\mathcal{P}_n^+(i,d,m)$.*

4.2 Network ψ-dependence

This section introduces a helpful framework to characterize the levels of network dependence between the regressors and the errors in equation (1). Form the vector $\mathbf{r}_{i,N} = [\mathbf{x}_{i,N}^\top, \varepsilon_{i,N}]^\top \in \mathbb{R}^{Q+1}$. For $Q, a \in \mathbb{N}_+$, endow $\mathbb{R}^{(Q+1)\times a}$ with the distance measure $\boldsymbol{d}_a(\mathbf{x},\mathbf{y}) = \sum_{l=1}^a \|x_l - y_l\|_2$ where $\|\cdot\|_2$ denotes the Euclidean norm and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{(Q+1)\times a}$. Let $\mathscr{L}_{Q,a}$ denote the collection of bounded Lipschitz real functions mapping values from $\mathbb{R}^{(Q+1)\times a}$ to $\mathbb{R}$. For each set of nodes A, let $\mathbf{r}_{A,N} = (\mathbf{r}_{N,i})_{i \in A}$. The ψ-dependence definitions in Doukhan and Louhichi (1999) and Kojevnikov et al. (2020) is accommodated to the proposed multilayer network framework.

Definition 2 (ψ-dependence) *A triangular array $\mathbf{r}_{n,i}$ for $n \geq 1$ and $\mathbf{r}_{n,i} \in n \in \mathbb{R}^{Q+1}$ is ψ-dependent if for each $n \in \mathbb{N}$ there is a sequence $\theta_n = \{\theta_{n,d}\}_{d \geq 0}, \theta_{n,0} = 1$ and a collection of non-random functions*

$(\psi_{a,b})_{a,b\in\mathbb{N}}, \psi_{a,b} : \mathscr{L}_{v,a} \times \mathscr{L}_{v,b} \to [0,\infty)$ *such that for all* $A, B \in \mathcal{P}_N^+(a,b,d)$ *for* $d > 0$ *and all* $f \in \mathscr{L}_{Q+1,a}$ *and* $g \in \mathscr{L}_{Q+1,b}$,

$$|\mathrm{Cov}\left(f\left(\mathbf{r}_{n,A}\right), g\left(\mathbf{r}_{n,B}\right)\right)| \leq \psi_{a,b}(f,g)\theta_{n,d}.$$

The sequence θ_n is called the dependence coefficients of $\mathbf{r}_{n,i}$. I state the definition in terms of triangular arrays because it fits the asymptotic results for the estimator in section 5. Given that the definition applies for any $n \in \mathbb{N}$, it can also be used to describe the dependence of the random variables in the population. Note that by choosing appropriate functions f and g, and appropriate sets A and B, Definition 2 bounds the covariance between any pair $\varepsilon_{N,i}$ and $\mathbf{x}_{N,j}$ (and also between any two $\mathbf{x}_{N,i}$ and $\mathbf{x}_{N,j}$). The following assumption guarantees that the dependence between individuals indeed decreases with their distance in the multilayer network space.

Assumption 2 (Weak Neighborhood Dependence (WND)) *Consider the set* $\mathscr{M}$ *of all possible realizations of* $\mathcal{M}_N$ *with positive probability mass in* $\mathcal{F}$. *For all networks* $\mathcal{M}_N \in \mathscr{M}$, *the conditional distribution* $\mathcal{F}(\mathbf{X}_N, \boldsymbol{\varepsilon}_{\boldsymbol{N}} \mid \mathcal{M}_N)$ *is such that:*

(i) $\{\mathbf{r}_{N,i}\}$ *is* ψ*-dependent with dependence coefficients* θ_N.

(ii) For $C > 0$, $\psi_{a,b}(f,g) \leq C \times ab\left(\|f\|_\infty + \mathrm{Lip}(f)\right)\left(\|g\|_\infty + \mathrm{Lip}(g)\right)$.

(iii) $\max_{d\geq 1}\theta_{N,d} < \infty$ *and there exists a finite constant* $D \in n \in \mathbb{N}_+$ *such that if* $d > D$, $\theta_{N,d} = 0$.

The WND Assumption is crucial for both identification and estimation. It is important to emphasize that the WND assumption holds for *any* network that happens with positive probability in the joint distribution $\mathcal{F}$. Therefore, even when the arbitrary network $\mathcal{M}_N$ forms endogenously, the conditions in WND holds for the conditional distribution $\mathcal{F}(\mathbf{X}_N, \boldsymbol{\varepsilon}_{\boldsymbol{N}} \mid \mathcal{M}_N)$. The idea is that any endogenous process of network formation will not place correlated individuals close to each other. Proposition A.3 in Appendix A in the Online Appendix presents a network formation process under which the WND assumption holds. Part *(ii)* of Assumption 2 states that the functional bounding the covariance in definition 2 is increasing in in the set sizes, the sup-norm of the aggregating functions f and g, and their Lipschitz constants, $\mathrm{Lip}(f)$ and $\mathrm{Lip}(g)$, related to the continuity of the functions.[1] Part *(iii)* imposes a stronger condition than Doukhan and Louhichi (1999) and Kojevnikov et al. (2020) because the dependence coefficients dissipate to zero after a finite distance D, not asymptotically. The existence of the finite distance D matters for identification. The reason is that the sharp bound allows forming identifying moment conditions based on exact distances (see Proposition 1). In addition to the weak dependence assumption, I shall impose an additional restriction related to the marginal distribution of the errors and their individual by individual correlation with the regressors.

[1] The Lipschitz constant for a function $f : \mathbb{R}^{(Q+1)\times a} \to \mathbb{R}$ is the smallest constant L such that$|f(\mathbf{x}) - f(\mathbf{y})| \leq C\boldsymbol{d}_a(\mathrm{x}, \mathrm{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{(Q+1)\times a}$.

Assumption 3 (Errors' Moments) *The unobserved shocks $\varepsilon_{N,i}$ are such that (i) $\mathbb{E}(\varepsilon_{N,i}) = 0$ and (ii) $\mathbb{E}(\mathbf{x}_{N,i}\varepsilon_{N,i}) = \mathbf{0}_{Q\times 1}$ for all $i \in \mathcal{I}_N$, where $\mathbf{0}_{Q\times 1}$ is the $Q \times 1$ vector of zeros.*

The first part of Assumption 3 can be viewed as just a normalization. Part *(ii)* is more substantial. It implies that the dependence between observed and unobserved characteristics is generated by the underlying network formation process but rules out any correlation between them for the same individual. This assumption is customary in network effects studies, see e.g., De Giorgi et al. (2020), Chan et al. (2022) and Zacchia (2019). Articles in the literature that are concern with endogenous network formation for the monolayer case use assumptions of no correlation between $\varepsilon_{N,i}$ and $\mathbf{x}_{N,i}$ after controlling for the effects of the underlying matching process. These type of estimators share the spirit of Assumption 3 in the sense that they think of the network formation process as generating the dependence between individuals' observed and unobserved attributes, see e.g, Auerbach (2016) and Johnsson and Moon (2019). Before presenting the proposition for the identifying condition, an additional measure of distance between two individuals i and j is introduced. The new distance is the second shortest path, and it is vital to guarantee that two individuals seemingly far away in terms of geodesic distance and number of edge type changes are not connected by a second path with larger geodesic distance but less edge type changes.

Definition 3 (Second Shortest Path) *Let $d^c_N(i,j)$ be the distance associated with the path connecting individuals i and j that has the second shortest number of links, and has a number of changes in edge types lower than $c^*_N(i,j)$. If no path exists with those characteristics, then $d^c_N(i,j) \to \infty$.*

For example, the second shortest path in Figure 1 for individuals 1 and 5 is given by $d_{1,5}(2) \to \infty$. That happens because there is not path connecting 1 and 5 that has less than two edge type changes. The following proposition characterizes the identifying moment conditions.

Proposition 1 *Let Assumptions 2 and 3 hold for $\mathbf{r}_{N,i}$. Then, there exist two constants K_c and K_d with $K_d > K_c + 1$ such that for all networks $\mathcal{M}_N \in \mathscr{M}$, the conditional distribution $\mathcal{F}(\mathbf{X}_N, \boldsymbol{\varepsilon}_N \mid \mathcal{M}_N)$ is such that for all pairs*

$$\mathbb{E}\left(\mathbf{x}_{N,j}\varepsilon_{N,i} \mid c^*_N(i,j) \geq K_c, d^c_N(i,j) \geq K_d\right) = \mathbf{0}_{Q\times 1}, \tag{6}$$

$$\mathbb{E}\left(\mathbf{x}_{N,j}\varepsilon_{N,i} \mid c^*_N(i,j) < K_c, d^*_N(i,j) \geq K_d\right) = \mathbf{0}_{Q\times 1}. \tag{7}$$

9 presents the proof for Proposition 1. The proof shows that if $K_d \leq K_c + 1$, the conditioning set in equation (6) does not provide different information from that in equation (7). Intuitively, the inequality $K_d > K_c + 1$ is required to guarantee that the dependence between individuals decreases faster when the number of edge type changes increase. Importantly, the proof sets $D = K_d$ and $\tau_{ij} =$

$\mathbb{1}\{d_N^c(i,j) \geq K_d\}K_c(K_d - K_c - 1)$, so that choosing the values of K_c and K_d completely characterizes the measure of distance $d_N^{\mathcal{M}}(i,j)$. The determination of the hyperparameters K_c and K_d is crucial for my identification argument. Their choice is best handled on a case-by-case basis, potentially involving theoretical arguments or re-sampling methods to justify the selection. Providing an optimal rule to choose those hyperparameters is an exercise left for future research.

Equation (6) can be interpreted in the context of the multilayer network structure. It states that paths involving edge types changes make the dependence between individuals decrease *faster*. Individuals connected by the same type of social or professional ties will tend to be similar because of the homophily characterizing the network formation process, see, e.g., Graham (2017). However, when additional types of connections are allowed, similar individuals connected by the same edge type can be indirectly connected to others who are different, given that they do not belong to that same *local* monolayer network. Equation (7) provides a result that has been used in recent literature for the case when $c_N^*(i,j) = 0$ for any i and j (monolayer network data case). It states that if non of the paths connecting two individuals change edge types enough times, they have to be at a longer distance in the network space for their characteristics to be not correlated. One illustrating case is when the network only includes intralayer edges. In that case, this assumption implies that if two individuals are too far apart *in the same layer*, it is more likely that their characteristics are *not* correlated. It is instructive to see how the examples presented before have implicitly used the results in Proposition 1.

Example 1 (continuation): In the context of non-overlapping network structure, De Giorgi et al. (2020) uses distance two coworkers of spouses and firms' shocks to identify peer effects in consumption.[2] Figure 1 helps to visualize the central identifying assumption. The idea is that shocks in firm 1, which are assumed to affect individuals' 1, 2, and 3 consumption levels, are independent of the unobserved characteristics of individual 5 who is indirectly connected to 3 by a coworker of a spouse relationship (see Section 4.1 in De Giorgi et al., 2020, pp. 142-144). In the context of Proposition 1, this means that it is enough to chose the parameter K_c to equal one in order to guarantee that equation (6) is satisfied when two types of connections are observed ($M = 2$). In general, the non-overlapping network structure guarantees that whenever it is possible to connect individuals i and j with a path containing at least one edge change, their interlayer distance is such that $d_N^c(i,j) \to \infty$. From Figure 1 it is clear that if $K_c = 1$ and K_d is arbitrarily large, it is still possible to find pairs such as $(2,4)$, $(1,4)$ or $(4,6)$ for which equation (6) holds. This property makes non-overlapping networks structures to be particularly useful for identification of peer effects, even more when it is combined with credible exogenous shocks as in De Giorgi et al. (2020).

Example 2 (continuation): Zacchia (2019) is concerned with the identification of contextual effects

[2] In Section 4.1 in (De Giorgi et al., 2020, pp. 142-144), the authors consider distance-3 nodes to be the same as intransitive triads. However, if we assume that each connection's weight is one, the shortest path measure for individuals connected by an intransitive triad is 2.

rather than peer effects. His setting is one where the network structure and some of the characteristics in $\mathbf{x}_{N,i}$ are allowed to be endogenously determined. Observing a monolayer network structure, the author proposes a game theoretical framework to formalize the idea that observed and unobserved attributes are orthogonal for individuals who are far in the (monolayer) network space. Assumption 1 in Zacchia (2019) resembles the restrictions imposed by equation (7) in this paper (see Section 2.1 in Zacchia, 2019, pp. 1994).

4.3 Theoretical Background

The identification and estimation approaches proposed in this paper do not require explicit assumptions about the network formation process. In order to strengthen their validity, I outline a game-theoretical model that generates the outcome equation (1) and under which the WND Assumption and the moment conditions in Proposition 1 hold. As mentioned throughout the text, I refer the reader interested on the model details to the Online Appendix A. A stylized two-stage game is provided where individuals first form a multilayer network with incomplete information, and then play a Bayesian social interaction game with strategic complementarities as in Calvó-Armengol et al. (2009) and Blume et al. (2015). The result in Proposition A.2 presents a theoretical foundation for the MLiM model, and a theoretical argument for the endogeneity issue caused by the network formation. Furthermore, the result in Proposition A.3 provides a theoretical justification for the sharp bound in the WND assumption, which justifies the use of the conditional moments (6) and (7) as a source for identification.

4.4 Identification Result

The identification argument combines the rich information provided by the multilayer network structure with the intuition that the strength of the dependence between individuals decays with their distance in the multilayer network space, as suggested by Proposition 1. To form the moment conditions required for identification, I construct the $(N \times N)$ matrices $\mathcal{W}_{N,m,\beta} = [w_{N,m,\beta;i,j}]$ and $\mathcal{W}_{N,m,\delta} = [w_{N,m,\delta;i,j}]$, where $w_{N,m,\beta;i,j}, w_{N,m,\delta;i,j} \in [0,1]$ are weights that are different from zero if (6) or (7) are satisfied for individuals i and j for layer m, respectively, and $w_{N,m,\beta;i,j} = w_{N,m,\delta;i,j} = 0$ otherwise. The sum of weights across rows may or may not sum up to one. These conditions apply for some fixed values of K_c and K_d. The two matrices are indexed by m because the paths connecting nodes i and j are required to start with an edge type m representing social effects generated by that layer. From the definition of the moment condition matrices, it follows that different individuals may provide different identifying power for the parameters of interest. For instance, if $\|\mathbf{w}_{N,m,\lambda,i}\|_1 = 0$ (where $\mathbf{w}_{N,m,\lambda,i}$ represents the ith row of $\mathcal{W}_{N,m,\lambda}$ and $\|\cdot\|_1$ represent the L_1 norm) then, individual i does not provide any information to identify the parameter λ, for $\lambda \in \{\beta, \delta\}$. In addition, isolated individuals do not contribute with information to identify the network effects parameters in this context.

To formalize this idea, notice that the joint distribution $\mathcal{F}$ induces a marginal distribution on the multilayer network $\mathcal{M}_N$, which determines the subset of individuals providing variation to identify the parameters of interest. Let the random variable $\eta_{N,m,i}$ equal one if individual i is non-isolated in layer m, and zero otherwise. Note that $\kappa_{N,m,i} = \mathbb{E}[\eta_{N,m,i}]$ gives the unconditional probability that individual i is non-isolated in layer m, where the expectation is taken with respect to the marginal distribution of $\mathcal{M}_N$. Additionally, it is possible to construct a set of random variables that contain possible measures of the expected identifying power for each individual i. These measures are based on the probability distribution of $\mathcal{W}_{N,m,\beta}$ and $\mathcal{W}_{N,m,\delta}$ induced by the marginal distribution of $\mathcal{M}_N$. Let $\eta_{N,m,\lambda,i}$ equals one if $\|\mathbf{w}_{N,m,\lambda,i}\|_1 > 0$ and zero otherwise. The expectation $\kappa_{N,m,\lambda,i} = \mathbb{E}[\eta_{N,m,\lambda,i}]$ represents the unconditional probability that individual i provides information to identify the parameter λ, where the expectation is taken with respect to the marginal distribution of $\mathbf{w}_{N,m,\lambda,i}$. Thus, $\kappa_{N,m,\lambda,i}$ measures the expected identifying power of individual i for any values of $\mathbf{X}_N$ and $\boldsymbol{\varepsilon}_N$ in their respective support. Let $\boldsymbol{\eta}_{N,\lambda,i} = [\eta_{N,1,\beta,i}, \ldots, \eta_{N,M,\beta,i}, \ldots, \eta_{N,M,\delta,i}]$ represent the random vector that collects all the $\eta_{N,m,\lambda,i}$ ($\boldsymbol{\kappa}_{N,\lambda,i}$ is defined analogously). The following assumption imposes restrictions on the random variables $\eta_{N,m,i}$ and $\eta_{N,m,\lambda,i}$ to guarantee that any configuration of the multilayer network that happens with positive provability provides enough information to identify the parameters of interest.

Assumption 4 (Identifying Variation) *For all $\mathcal{M}_N \in \mathscr{M}$ and associated marginal distributions, the random variables $\eta_{N,m,i}$ and $\eta_{N,m,\lambda,i}$ are such that: (i) The event $\eta_{N,m,i} = 0$ for all i and m, happens with probability zero. (ii) The event $\eta_{N,m,\lambda,i} = 0$ for all i, m and λ, happens with probability zero.*

Parts (i) and (ii) are crucial for identification. If $\eta_{N,m,i} = 0$ for all i and m, then all individuals are isolated in all layers and the population multilayer network would not provide any information about social interactions that can be used for identification. Similarly, even if not all individuals are isolated, but $\eta_{N,m,\lambda,i} = 0$ for all i, m and λ, then it is not possible to form moment conditions to back up the parameters of interest.

Identification of the model in (1) requires at least $(M+1)(Q+1)$ moment conditions, given that there are $(M+1)(Q+1)$ parameters to estimate. If the $2M$ matrices $\mathcal{W}_{N,m,\beta}$ and $\mathcal{W}_{N,m,\delta}$ provide different information for each m, in the sense that their expectations are linearly independent, then it is possible to construct at least $(2M+1)Q+1$ moment conditions based on the Q observable characteristics $\mathbf{x}_{N,i}$. Define the matrix $\mathbf{D}_N = [\mathbf{W}_{N,1}\mathbf{y}_N, \ldots, \mathbf{W}_{N,M}\mathbf{y}_N, \mathbf{W}_{N,1}\mathbf{X}_N, \ldots, \mathbf{W}_{N,M}\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$ associated with the vector $\boldsymbol{\psi} = [\beta_1, \ldots, \beta_M, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_M, \widetilde{\boldsymbol{\gamma}}]$, where $\widetilde{\mathbf{X}}_N = [\boldsymbol{\iota}_N, \mathbf{X}_N]$, and $\widetilde{\boldsymbol{\gamma}} = [\alpha, \boldsymbol{\gamma}]$. As before, the true parameters are denoted by $\boldsymbol{\psi}^0 = [\beta_1^0, \ldots, \beta_M^0, \boldsymbol{\delta}_1^0, \ldots, \boldsymbol{\delta}_M^0, \alpha^0, \boldsymbol{\gamma}^0]$. For any given K_c and K_d, the matrices $\mathcal{W}_{N,m,\beta}$ and $\mathcal{W}_{N,m,\delta}$ can be used to construct the matrix associated with the moment conditions given by $\mathbf{Z}_N = [\mathcal{W}_{N,1,\beta}\mathbf{X}_N, \ldots, \mathcal{W}_{N,M,\beta}\mathbf{X}_N, \mathcal{W}_{N,1,\delta}\mathbf{X}_N, \ldots, \mathcal{W}_{N,M,\delta}\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$. The previously defined matrices $\mathbf{D}_N$ and $\mathbf{Z}_N$ have dimensions $N \times (M+1)(Q+1)$ and $N \times (1+R+Q)$, respectively. Note that $(M+1)(Q+1) = 1 + M + QM + Q$, so that $R > M + QM$ when there are more than one regressor in matrix $\mathbf{X}_N$ that can be used as an instrument for the peer effects variables. The vector $\boldsymbol{\psi}$ has dimensions $(M+1)(Q+1) \times 1$, so that depending on the value of Q the system can be over or just identified.

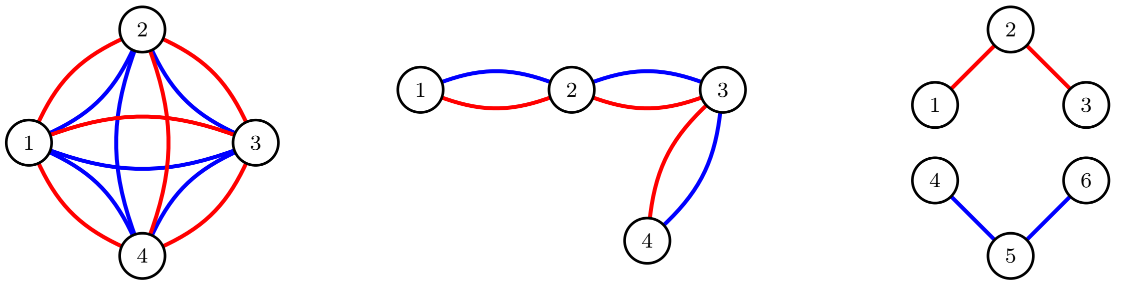
Before formalizing the identification argument, it is relevant to include a remark regarding the heterogeneity in identifying power among different individuals in the population. Let $\mathbf{z}_{N,i}$ and $\mathbf{d}_{N,i}$ represent the ith row of the matrix $\mathbf{Z}_N$ and $\mathbf{D}_N$, respectively. Also, define the $(1+R+Q)\times(1+R+Q)$ matrix matrix $\mathcal{H}_{N,\lambda,i} = \text{diag}(\eta_{N,1,\beta,i}\boldsymbol{\iota}_Q, \ldots, \eta_{N,M,\delta,i}\boldsymbol{\iota}_Q, \boldsymbol{\iota}_{Q+1})$, with $\mathcal{K}_{N,\lambda,i} = \mathbb{E}[\mathcal{H}_{N,\lambda,i}]$. By Proposition 1, and the definition of $\mathbf{z}_{N,i}$, it follows that $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})] = \mathbf{0}_{1+R+Q}$, and by the law of total expectation, the expression can be written as $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})] = \mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi}) \mid \mathcal{H}^*_{N,\lambda,i} \neq \mathbf{O}_{R\times R}] + (\mathcal{I}_{1+R+Q} - \mathcal{K}_{N,\lambda,i})\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi}) \mid \mathcal{H}^*_{N,\lambda,i} = \mathbf{O}_{R\times R}]$, where $\mathcal{H}^*_{N,\lambda,i}$ contains the left top upper matrix of $\mathcal{H}_{N,\lambda,i}$, and $\mathbf{O}_{R\times R}$ is the $(R\times R)$ matrix of zeros. Note that when $\mathcal{H}^*_{N,\lambda,i} = \mathbf{O}_{R\times R}$ the upper left matrix of the conditional expectation is trivially zero, and the $(Q+1)$ lower right component of $(\mathcal{I}_{1+R+Q} - \mathcal{K}_{N,\lambda,i})$ is also a matrix of zeros. Therefore, $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})] = \mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi}) \mid \mathcal{H}^*_{N,\lambda,i} \neq \mathbf{O}_{R\times R}]$. Defining the moment condition function as $\mathbf{m}_N(\boldsymbol{\psi}) = \sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})$, it follows that $\mathbb{E}[\mathbf{m}_N(\boldsymbol{\psi})]$ can be written as $\sum_{i\in\mathcal{I}_N}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})] = \sum_{i\in\mathcal{I}_N}\mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi}) \mid \mathcal{H}^*_{N,\lambda,i} \neq \mathbf{O}_{R\times R}]$. Intuitively, the moment condition is a weighted sum of conditional expectations where the weights are the unconditional probability that $\mu_{N,m,\lambda,i} = 1$. This weighting scheme gives more importance to individuals for whom the probability of finding moment conditions is higher in the population.

For identification to be possible, in addition to the restrictions guaranteeing the validity of the moment conditions, the instrumental variables in $\mathbf{z}_{N,i}$ need to be relevant. The relevance condition is closely related with the identifying variation, as the second is a necessary condition (but not sufficient) for the first. To see why, notice that by the same arguments presented before, it follows that $\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}^\top_{N,i}] = \mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}^\top_{N,i} \mid \mathcal{H}^*_{N,\lambda,i}, \mathcal{H}^*_{N,i} \neq \mathbf{O}_{R\times R}]\mathcal{K}_{N,i}$, where $\mathcal{K}_{N,i}$ is a $(M+1)(Q+1)\times(M+1)(Q+1)$ matrix such that $\mathcal{K}_{N,i} = \mathbb{E}[\mathcal{H}_{N,i}]$, for $\mathcal{H}_{N,i} = \text{diag}(\eta_{N,1,i}\boldsymbol{\iota}_Q, \ldots, \eta_{N,M,i}\boldsymbol{\iota}_Q, \boldsymbol{\iota}_{Q+1})$, and $\mathcal{H}^*_{N,i}$ is defined analogously to $\mathcal{H}^*_{N,\lambda,i}$. Therefore, if $\eta_{N,m,i} = \eta_{N,m,\lambda,i} = 0$ for all i and m and λ, the matrix $\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\mathbf{d}^\top_{N,i}]$ cannot have full column rank. The following assumption imposes a column rank condition on the matrix $\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\mathbf{d}^\top_{N,i}]$.

Assumption 5 (Relevance) *The matrix* $\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\mathbf{d}^\top_{N,i}] < \infty$ *has full column rank.*

As mentioned above, parts *(i)* and *(ii)* of Assumption 4 are necessary conditions for Assumption 5 to hold. If either all individuals are isolated, or it is not possible to form moment conditions for any i, then the full rank condition is trivially not satisfied. Additionally, the relevance assumption has empirical consequences. For n large enough, the matrix $\mathbf{Z}^\top_n\mathbf{D}_n$ has to have full column rank, for which a necessary condition is that the matrices $\mathbf{W}_{n,1}, \ldots, \mathbf{W}_{n,M}$ and $\mathbf{I}_n$ are linearly independent. This imposes restrictions on the m adjacency matrices for the layers in network $\mathcal{M}_n$, they have to be all different from each other, and non-zero. Moreover, for the instruments in $\mathbf{Z}_n$ to not be redundant, one necessary condition is that the matrices $\mathcal{W}_{n,1,\beta}$, $\mathcal{W}_{n,2,\beta}$, $\ldots$, $\mathcal{W}_{n,M,\beta}$, $\mathcal{W}_{n,1,\delta}$, $\mathcal{W}_{n,2,\delta}$, $\ldots$, $\mathcal{W}_{n,M,\delta}$ and $\mathbf{I}_n$ are linearly independent.

Figure 3: Failure of Assumption 5



Panel (a) Panel (b) Panel (c)

Note: Panels (a), (b) and (c) display examples of flat representations of different multilayer networks for $M = 2$. In panel (a) the multilayer network is dense, meaning that all individuals are connected to each other in both layers. The multilayer network in panel (b) contains intransitive triads in both individual layers, and the adjacency matrices of both layers are equal. Panel (c) shows a multilayer network with intransitive triads in both layers, and linearly independent adjacency matrices for each layer. The number of edge type changes for the three multilayer networks equals zero for all possible shortest paths. Thus, for the presented examples, $\mathcal{W}_{n,m,\beta} = \mathbf{O}$ for all $m \in \{1, 2\}$ and $n = 4$ or $n = 6$ (where $\mathbf{O}$ represents the matrix of zeros). For the example in panel (a), for any $K_d > 1$, $\mathcal{W}_{m,\delta} = \mathbf{O}$ for all $m \in \{1, 2\}$, as the minimum path length for any nodes i and j is always 1.

These conditions are not trivially satisfied by $\mathcal{W}_{n,m,\beta}$ and $\mathcal{W}_{n,m,\delta}$ as there may not be enough (or too much) sparsity in the intralayer and interlayer connections to guarantee that Assumption 5 holds for given values of K_c with K_d. Figure 3 presents three examples where Assumption 5 fails. Panels (a), (b) and (c) illustrate flat representations of different multilayer networks for $M = 2$. This example's main feature is that the number of edge type changes is zero for all the presented multilayer networks in all possible shortest paths between two individuals i and j, for all $(i, j) \in V_m$, where $m = \{1, 2\}$. Having zero changes in edge types for all shortest paths implies that $\mathcal{W}_{n,m,\beta} = \mathbf{O}$ for all $m \in \{1, 2\}$ and $n = 4$ (where $\mathbf{O}$ represents the matrix of zeros). Assumption 5 breaks down for the examples in panels (a) and (b) because both layers' adjacency matrices are equal. Therefore, the necessary condition of linear independence of the m layers' adjacency matrices $\mathbf{W}_{n,m}$ is not satisfied. Panel (c) presents an example where the two layers' adjacency matrices are different and linearly independent, but still, Assumption 5 fails. The reason is that the two networks are completely disjointed. These examples show that linear independence is only a necessary condition and that the layers also need to have connections in common for Assumption 5 to be satisfied. Notably, as shown in the examples presented in panels (b) and (c), unlike previous work in the monolayer case, intransitive triads in each layer's network are not enough to guarantee identification. See Bramoullé et al. (2009) and De Giorgi et al. (2010) for seminal work on the importance of intransitive triads for identification in the monolayer case.

Instruments' relevance has an additional interpretation when using a series expansion on the solution of (2). The conditions in Assumption 1 required for invertibility also guarantee that each $\mathbf{W}_{N,m}\mathbf{y}_N$ can be written as a function of infinite powers of $\mathbf{W}_{N,m}$ and infinite products of $\mathbf{W}_{N,m}\mathbf{W}_{N,s}$ for all $m = 1, \ldots, M$ and $m \neq s$ (see Remark 2 in Online Appendix B for more details). In the monolayer network case, it is well known that the (i, j)th element of the matrix $\mathbf{W}^k_{N,m}$ gives the number of paths of length k from agents i to j (for some layer m), see e.g., Graham (2015). For the multilayer case, assuming that $V_m = V$ for all m, the (i, j)th element of the product of two adjacency matrices $\mathbf{W}_{N,m}$ and $\mathbf{W}_{N,s}$ for layers m and s, contains the number of paths of length two between nodes i and j, where each path begins with a type m edge and changes to type s after the second node in the sequence. Remark 2 in Appendix B shows that both interlayer and intralayer indirect connections can be used as relevant instruments for $\mathbf{W}_{N,m}\mathbf{y}_N$ given some conditions on the parameters. The rule for choosing which indirect connections are also valid is given in Proposition 1. The following theorem formalizes the previous discussion.

Theorem 1 *Suppose Assumptions 1, 2, 3, 4, and 5 hold for some K_c and K_d such that $K_d > K_c + 1$. Then, the parameters $\boldsymbol{\psi}^0 = [\beta^0_1, \ldots, \beta^0_M, \boldsymbol{\delta}^0_1, \ldots, \boldsymbol{\delta}^0_M, \alpha^0, \boldsymbol{\gamma}^0]$ are identified by the moment conditions $\mathbb{E}[\mathbf{m}_N(\boldsymbol{\psi})] = 0$, where $\mathbf{m}_N(\boldsymbol{\psi}) = \sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})$.*

Remark 1 Theorem 1 is based on the existence of a set of regressors for which Assumption 3 applies. Note that only one of such regressors is necessary to form sufficient moment conditions to identify $\beta_1, \ldots, \beta_m$. The parameters $\boldsymbol{\gamma}$ are only identified for the set of regressors for which part *(ii)* of Assumption 3 is satisfied. It is possible to control for other observable characteristics that are not orthogonal to the errors as long as they are uncorrelated with the set of exogenous regressors. The estimated parameters for those regressors do not have a causal interpretation. Identification of $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_M$ requires the existence of at least Q exogenous regressors. The validity of the instruments formed by the $\mathcal{W}_{m,\delta}$ matrices is guaranteed by Proposition 1. The strength of those instruments, though, should be considered on a case-by-case basis, see, e.g., Zacchia (2019).

9 presents the proof for Theorem 1. This theorem shows that identification is possible in a general framework that allows for endogenous multilayer network formation and potential correlation between the regressors and errors. This result exploits the additional information provided by the multilayer network structure and depends crucially on the possibility of forming valid paths. The level of generality of this identification approach makes it applicable to different models and data structures proposed in the literature. As to showcase the generality of Theorem 1, the following paragraphs describe how it can be applied in the context of the two examples presented in the previous sections.

Example 1 (continuation): Given their data structure, De Giorgi et al. (2010) rules out the possibility of peer effects generated by the spouse's network. Assuming that the social effects from the spouses' network equal zero, and ignoring the panel data structure for simplicity, the linear model (1) reduces to

$$y_{N,i} = \alpha + \sum_{j \neq i} w_{N,1;i,j} y_j \beta + \sum_{j \neq i} w_{1;i,j} \mathbf{x}_{N,j}^\top \boldsymbol{\delta} + \mathbf{x}_{N,i}^\top \boldsymbol{\gamma} + \varepsilon_{N,i}, \tag{8}$$

where $\mathbf{W}_{N,1}$ is the co-workers network of interest. The researcher also observes the network of spouses $\mathbf{W}_{N,2}$. De Giorgi et al. (2020) assumes that $K_c = 1$, and the structure of non-overlapping networks guarantee that equation (6) in Proposition 1 holds for any K_d. With this information, it is possible to construct $\mathcal{W}_{N,1,\beta}$ which can be used to identify the co-workers peer effects β. By choosing $K_c = 1$ and any arbitrary K_d, it is possible to construct $\mathcal{W}_{N,1,\beta}$ for the example in Figure 1:

$$\mathcal{W}_{N,1,\beta} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array}.$$

De Giorgi et al. (2020) uses individuals connected by length-two paths that change their edge type from co-workers to spouses as an instrument for the peer effects endogenous variable. This set of paths can be characterized by the matrix product of the two layers' adjacency matrices $\mathbf{W}_{N,1}\mathbf{W}_{N,2}$, and it is a subset of the valid paths presented in the matrix $\mathcal{W}_{N,1,\beta}$. Assuming only one regressor for the sake of illustrations, note that $\mathbb{E}[\mathcal{W}_{N,1,\beta}\mathbf{x}_N\boldsymbol{\varepsilon}_N]$ produces a vector of expectations with components such as $\mathbb{E}[x_4\varepsilon_1]$, $\mathbb{E}[x_5\varepsilon_1]$, $\mathbb{E}[x_6\varepsilon_1]$, $\mathbb{E}[x_7\varepsilon_1]$, which are all equal to zero under the results in Proposition 1 for $K_c = 1$, and where $d^1(i,j) \to \infty$ follows from the network structure for all i and j.

De Giorgi et al. (2020) does not use any instrument to identify contextual effects. In light of Proposition 1, it is possible to understand the absence of an instrument for contextual effects as assuming $K_d = 0$. Under these assumptions, the matrix $\mathcal{W}_{N,1,\delta}$ reduces to $\mathcal{W}_{N,1,\delta} = \mathbf{W}_{N,1}$, and the matrix of instruments is given by $\mathbf{Z}_N = [\mathbf{W}_{N,1}\mathbf{W}_{N,2}\mathbf{X}_N, \mathbf{W}_{N,1}\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$. Theorem 1 can be apply to show that $[\beta, \boldsymbol{\delta}, \alpha, \boldsymbol{\gamma}]$ are indeed point identified.

Example 2 (continuation): Zacchia (2019) considers a model where only contextual effects are relevant, and the researcher observes only one network. His model can be written as

$$y_{N,i} = \alpha + \sum_{j \neq i} w_{N,i,j} \mathbf{x}_{N,j}^\top \boldsymbol{\delta} + \mathbf{x}_{N,i}^\top \boldsymbol{\gamma} + \varepsilon_{N,i},$$

where δ is the coefficient of interest for Zacchia (2019). The author has access to panel data but, for simplicity, this example considers the cross-section case. Because the author is assuming that the peer effects are zero, the only relevant matrix in this case is $\mathcal{W}_{N,\delta}$ constructed based on equation 7. Zacchia (2019) chooses $K_d = 2$ but also includes empirical estimations for $K_d = 3$. Assuming the parameter value $K_d = 2$, the matrix to identify contextual effects for the network in Figure 2, this matrix is given by

$$\mathcal{W}_{n,\delta} = \begin{array}{c} \\ i \\ j \\ k \\ l \end{array} \begin{array}{c} \begin{array}{cccc} i & j & k & l \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}.$$

Based on $\mathcal{W}_{N,\delta}$, the matrix of instruments can be constructed as $\mathbf{Z}_N = [\mathcal{W}_{N,\delta}\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$, and $\mathbf{D}_N = [\mathbf{W}_N\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$. Theorem 1 can be also apply for this case to show that the parameters $[\boldsymbol{\delta}, \alpha, \boldsymbol{\gamma}]$ are point identified.

5 Estimation

This section provides details for the construction of a GMM estimator for the vector of parameters $\boldsymbol{\psi}^0$. The empirical counterparts of the moment conditions used for identification in Theorem 1 are the basis for estimation. Consider the joint distribution $\mathcal{F}$ that characterizes the values of $\mathbf{x}_{N,i}$, $\varepsilon_{N,i}$, and $\mathcal{M}_N$ in an arbitrarily large population where outcomes follow the model in equation (1). Assume that a practitioner observes a random sample of size $n < N$ from that population that preserves the network structure in $\mathcal{M}_N$. This article interprets the sampling mechanism as in Graham (2020). The sample schema is a thought experiment useful to derive limiting distributions convenient in practice.

From the sample, the analyst observes $\{y_i, \mathbf{x}_i^\top, \{\{w_{m;i,j}\}_{j=1,j\neq i}^n\}_{m=1}^M\}_{i=1}^n$ (or $\{\mathbf{y}_n, \mathbf{X}_n, \mathcal{M}_n\}$ in vector form). Then, it is possible to construct the $n \times (M+1)(Q+1)$ matrix of regressors $\mathbf{D}_n$, and the $n \times (1 + R + Q)$ matrix of instruments $\mathbf{Z}_n$. The errors are a function of the unknown parameters, $\boldsymbol{\varepsilon}_n(\boldsymbol{\psi}) = \mathbf{y}_n - \mathbf{D}_n\psi$ according to equation (1). The construction of the matrix of instrument $\mathbf{Z}_n$ requires the computation of the matrices $\mathcal{W}_{n,m,\beta}$ and $\mathcal{W}_{n,m,\delta}$ for $m = 1\ldots, M$, which involves evaluating equations (6) and (7) for all the possible dyads $i, j \in \mathcal{I}_n \times \mathcal{I}_n$ (where $\mathcal{I}_n$ represents the set of individuals in the sample). This problem could entail a computational complexity in the order of $n^{\binom{n}{2}}$. I use two algorithms that are based on the idea behind Dijkstra's shortest path algorithm for monolayer graphs. The modified algorithms compute the first and second shortest paths with a minimum number of edge types in polynomial time (Balasubramanian et al., 2022). The algorithms are described in the Online Appendix C and may be of independent interest for those interested in path problems.

The GMM estimator is formed based on the moments $J_N(\boldsymbol{\psi}) = \mathbb{E}[\mathbf{m}_N(\boldsymbol{\psi})]^\top(\mathbf{A}_N^\top\mathbf{A}_N)\mathbb{E}[\mathbf{m}_N(\boldsymbol{\psi})]$, with sample analog given by $J_n(\boldsymbol{\psi}) = [n^{-1}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n(\boldsymbol{\psi})]^\top(\mathbf{A}_n^\top\mathbf{A}_n)[\mathbf{Z}_N^\top\boldsymbol{\varepsilon}_n(\boldsymbol{\psi})]$, where $\mathbf{A}_n$ is a fixed $(M+1)(Q+1)\times(1+R+Q)$ full row-rank matrix, assumed to converge to a constant full row-rank matrix $\mathbf{A}_N$. The linearity assumption of the model in (1) guarantees that the GMM estimator has a closed form given by

$$\widehat{\boldsymbol{\psi}}_{GMM} = [\mathbf{D}_n^\top\mathbf{Z}_n(\mathbf{A}_n^\top\mathbf{A}_n)\mathbf{Z}_n^\top\mathbf{D}_n]^{-1}[\mathbf{D}_n^\top\mathbf{Z}_n(\mathbf{A}_n^\top\mathbf{A}_n)\mathbf{Z}_n^\top\mathbf{y}_n].$$

The GMM estimator $\widehat{\boldsymbol{\psi}}_{GMM}$ is constructed based on a random sample from the joint distribution $\mathcal{F}(\mathbf{X}_N, \mathcal{M}_N, \boldsymbol{\varepsilon}_N)$ which allows for correlation between the errors, the regressors and the multilayer network. As discussed in section 4.1, the week dependence Assumption in 2 controls the levels of dependence between individuals based on their distance in the multilayer network space. Given that the data used to calculate $\widehat{\boldsymbol{\psi}}_{GMM}$ are assumed to come from a random sample of a week dependent population, they will inherit that property for all finite samples of size n. In addition to the week dependence condition, the following assumptions are necessary to characterize the asymptotic behavior of the linear GMM estimator in a context where network dependence is allowed.

Assumption 6 (Existence of Moments) *Let the functions $f_{q,\ell}$ and $g_{q',\ell'}$ mapping $\mathbb{R}^{(Q+1)\times 2}$ to $\mathbb{R}$ be such that $f_{q,\ell}(\mathbf{r}_{n,\{i,j\}}) = r_{n,i}^q r_{n,j}^\ell$ and $g_{q',\ell'}(\mathbf{r}_{n,\{h,s\}}) = r_{n,h}^{q'} r_{n,s}^{\ell'}$ for $i,j,h,s \in \mathcal{I}_n$, $i \neq j$, $h \neq s$, $q \neq \ell$ and $q' \neq \ell'$. Assume that $\|f_{q,\ell}(\mathbf{r}_{n,\{i,j\}})\|_{p_f^*} + \|g_{q',\ell'}(\mathbf{r}_{n,\{h,s\}})\|_{p_g^*} < \infty$ for all q,ℓ where $p_f^* = \max\{p_{f,i}, p_{f,j}\}$ (analogous for p_g^*) and $1/p_{f,i} + 1/p_{f,j} + 1/p_{g,h} + 1/p_{g,s} < 1$.*

This assumption imposes the existence of moments for non-linear functions of ψ-dependent random variables. The existence of these moments is required to guarantee that the covariances between the transformed random variables for two groups of individuals are bounded for a large enough distance in the network space. Importantly, this assumption does not impose any differentiating restrictions on the correlation structure between the regressors and the errors than in the autocorrelations between two sets of regressors. The following Assumption guarantees that the sum of the weights for any adjacency matrix $\mathbf{W}_{n,m}$, and any matrix of moment conditions $\mathcal{W}_{n,\lambda,m}$ does not grow faster than the sample size. Similar to the notation introduced in Definition 1, let $\mathcal{P}_n(i,1,m,\lambda) = \{i : i,j \in \mathcal{I}_N\text{, and } w_{n,m\lambda;i,j} > 0\}$.

Assumption 7 (Bounded Weights) *For all networks $\mathcal{M}_n \in \mathscr{M}$, (i) for all layers m, coefficient λ, and all individuals $i \in \mathcal{I}_n$, $\sum_{j\in\mathcal{P}_n(i,1,m,\lambda)} w_{n,m,\lambda;i,j} = o_{a.s.}(n)$ and $\sum_{j\in\mathcal{P}_n(i,1,m)} w_{n,m;i,j} = o_{a.s.}(n)$. (ii) the set of individuals formed by the intersection of non-isolated nodes in layer m and $\lambda \in \{\beta,\delta\}$ denoted by $\nu_{m,\lambda}$, and those from the the product of ζ adjacency matrices organized in a sequence ϕ, denoted by $\nu_{\zeta,\phi,\lambda,m_1}$, is such that $\sum_{j\in\eta_{i,\mu,\phi}} w_{\mu,\phi;i,j} = o_{a.s.}(n)$ for any $j \in \mathcal{I}_n$, all $(i,s) \in \eta_{\mu,\phi}$, any product of adjacency matrices ζ and any sequence ϕ.*

Part (i) of Assumption 7 guarantees that the weighted sum of individual connections and nodes that can be used as instruments do not grow faster than the sample size. This condition is generally satisfied. In particular, it is immediately satisfied when the matrix $\mathcal{W}_{n,m,\lambda}$ and $\mathcal{W}_{n,m}$ are row-normalized for $\lambda \in \{\beta, \delta\}$ and any m. The reason is that $\sum_{j \in \mathcal{P}_n(i,m,\lambda,1)} w_{m,\lambda;i,j} = 1$ for any individual i (and the same is true for the adjacency matrices of the layers). The use of row-normalized adjacency matrices is common in the literature of econometrics of networks, see, e.g., de Paula (2017). This assumption is related with the relevant condition for identification in section 3, where I pointed out that too dense or too sparse population networks can breakout identification. Part *(ii)* guarantees that the number of paths of order ζ does not grow faster than n. This is a technical requirement necessary for the moments of the outcome $\mathbf{y}_n$ to exist. The next assumption imposes a global measure of sparsity on the asymptotic multilayer network $\mathcal{M}_N$.

Assumption 8 (Dependence Rate of Decay) *Let $\bar{D}_n(d) = n^{-1} \sum_{i \in \mathcal{I}_n} |\mathcal{P}_n(i,d)|$ be the average number of distance-d connections on the multilayer network network $\mathcal{M}_n$. Then, for all networks $\mathcal{M}_N \in \mathscr{M}$, $n^{-1} \sum_{d \geq 1} \bar{D}_n(d) \theta_{n,d} \xrightarrow{a.s.} 0$ as $n \to \infty$.*

Assumption 8 captures the trade off between the network density and the level of dependence that the model can allow via the dependence coefficients $\theta_{n,d}$. Intuitively, it guarantees that, on average, the level of sparsity captured by $\bar{D}_n(d)$ does not increase faster than the dependence between individuals at distance d for all possible distances. Assumption 8 extends Assumption 3.2 in Kojevnikov et al. (2020) for the case where the distance between individuals is defined in the multilayer network space. In addition to characterizing the rate of decay for the dependence coefficients $\theta_{n,d}$, the central limit theorem result in Kojevnikov et al. (2020) requires to impose sparsity restrictions based on average neighborhood sizes and average neighborhood shell sizes. Following Kojevnikov et al. (2020), define a measure for the average neighborhood size as $\delta_n(d;k) = n^{-1} \sum_{i \in \mathcal{I}_n} |\mathcal{P}_n(i,d)|^k$, and a measure for the average neighborhood shell size as

$$\delta_n^{-}(d,m;k) = \frac{1}{n} \sum_{i \in \mathcal{I}_n} \max_{j \in \mathcal{P}_n(i,d)} \left| \mathcal{P}_n^{-}(i,m) \setminus \mathcal{P}_n^{-}(j,d-1) \right|^k ,$$

were $\mathcal{P}_n^{-}(j,d-1) =$ when $d = 0$. With these two measures of average density, construct the combined quantity

$$c_n(d,m,k) = \inf_{\alpha > 1} \left[\delta_n^{-}(d,m,k\alpha) \right]^{\frac{1}{\alpha}} \left[\delta_n \left(d; \frac{\alpha}{\alpha - 1} \right) \right]^{1 - \frac{1}{\alpha}} . \tag{9}$$

The measure of average density in equation 9 is crucial to impose a set of assumptions that are sufficient for Kojevnikov et al.'s (2020) central limit theorem to apply. For an arbitrary position q in $\mathbf{Z}_{n,i}$, define $S_n = \sum_{i \in \mathcal{I}_n} z_{n,i,q} \varepsilon_{n,i}$. Defining $\sigma_n^2 = \text{Var}(S_n)$, the following assumption guarantees the existence of higher order moments, imposes asymptotic sparsity, and bounds the long-run variance.

Assumption 9 (Average Sparsity) *For all networks $\mathcal{M}_N \in \mathscr{M}$,* *(i) for some $p > 4$, $\sup_{n\geq 1} \max_{i\in\mathcal{I}_n} \|z_{n,i,q}\varepsilon_{n,i}\|_p < \infty$.*

There exists a sequence $m_n \to \infty$ such that for $k = 1, 2$,

(ii) , $\frac{n_q}{\sigma_n^{2+k}} \sum_{d\geq 0} c_n(d, m_n, k)\, \theta_{n,d}^{1-\frac{2+k}{p}} \xrightarrow{a.s.} 0$ as $n \to \infty$,

(iii) $\frac{n^2 \theta_{n,m_n}^{1-(1/p)}}{\sigma_n} \xrightarrow{a.s.} 0$ as $n \to \infty$.

Lemmas 6 and 7 in the Online Appendix B show that the sample averages $n^{-1}\mathbf{D}_n^\top \mathbf{Z}_n$ and $n^{-1}\mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n$ converge to their population counterparts $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]$ and $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})]$, respectively. Moreover, Lemma 9 shows that $\boldsymbol{\Omega}_n = \mathrm{Var}(\mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n)$ converges to the population variance

$$\boldsymbol{\Omega}_N = \lim_{n\to\infty} n^{-1} \left[\sum_{i=1}^{n} \mathrm{var}(\mathbf{z}_{n,i}\varepsilon_{n,i}) + \sum_{i\neq j} \mathrm{cov}(\mathbf{z}_{n,i}\varepsilon_{n,i}, \mathbf{z}_{n,j}\varepsilon_{n,j}) \right] < \infty, \tag{10}$$

which is necessary for the multivariate central limit theorem result in Lemma 10. Summing over all individuals is analogous to summing over all possible distances. By definition, $\mathbb{E}[\mathbf{z}_{n,i}\varepsilon_{n,i}] = 0$ for all i and n. Then, the limiting measure of covariance in equation (10) can be written in terms of a generic distance d as

$$\boldsymbol{\Gamma}_N(d) = \sum_{i\in\mathcal{I}_N} \sum_{j\in\mathcal{P}_N(i,d)} \mathbb{E}\left[\mathbf{z}_{N,i}\varepsilon_{N,i}\varepsilon_{N,j}\mathbf{z}_{N,j}^\top\right], \tag{11}$$

which implies that the population variance-covaraince matrix $\boldsymbol{\Omega}_N$ can be constructed as the sum of the covariance estimators in equation (11) for all possible distances $d \geq 0$ (equality is allowed to account for the variance),

$$\boldsymbol{\Omega}_N = \sum_{d\geq 0} \boldsymbol{\Gamma}_N(d). \tag{12}$$

After characterizing the variance-covariance matrix $\boldsymbol{\Omega}_N$ both in terms of individual and distances sums, Theorem 2 shows that the optimal GMM estimator is consistent and asymptotically normal. The optimal estimator is given by $\widehat{\boldsymbol{\psi}}_{GMM}^\star = (\mathbf{D}_n^\top \mathbf{Z}_n \boldsymbol{\Omega}_N^{-1} \mathbf{Z}_n^\top \mathbf{D}_n)^{-1}(\mathbf{D}_n^\top \mathbf{Z}_n \boldsymbol{\Omega}_N^{-1} \mathbf{Z}_n^\top \mathbf{y}_n)$.

Theorem 2 *Let Assumptions 1-8 hold, then $\widehat{\boldsymbol{\psi}}_{GMM}^\star = \boldsymbol{\psi} + o_p(1)$ and $\sqrt{n}(\widehat{\boldsymbol{\psi}}_{GMM}^\star - \boldsymbol{\psi}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_N^\star)$, where $\boldsymbol{\Sigma}_N^\star = (\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]^\top \boldsymbol{\Omega}_N^{-1} \mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top])^{-1}$.*

9 presents the proof for Theorem 2. As discussed before, the expectation $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]$ can be written as $\sum_{i\in\mathcal{I}_N} \mathcal{K}_{N,\lambda,i} \mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top \mid \mathcal{H}_{N,\lambda,i}^*, \mathcal{H}_{N,i}^* \neq \mathbf{O}_{R\times R}]\mathcal{K}_{N,i}$, where $\mathcal{K}_{N,\lambda,i}$ and $\mathcal{K}_{N,i}$ contain the probabilities of finding moment conditions and the probability of being isolated for individual i in all possible layers m. These probabilities implicitly depend on the network formation model and are a function of the network's multilayer connectivity. The results in this theorem are interpreted as the network counterpart of the results by Lee (2007b) in the context of group structures. Lee (2007b) shows

that the network parameters of peer and contextual effects have a slower convergence rate than the direct effect coefficients and that they slow down when the groups' size increases. Intuitively, for larger groups, social interactions are more challenging, which could be interpreted as more isolated individuals reducing social interactions in the network framework. Instead of affecting the rate of convergence, this paper shows that, in the context of multilayer networks, the existence of isolated individuals directly affects the variance-covariance matrix of the peer and contextual effects estimators. Moreover, $\mathbf{\Sigma}_N^\star$ depends on the population probabilities of an individual providing identification information. When that probability approaches zero, the $2Q$ upper right sub-matrix of $\mathcal{K}_{N,\lambda,i}$ approaches the zero matrix, and the variance-covariance matrix could grow arbitrarily large. In the extreme case of non-identification, the matrix $\widehat{\boldsymbol{\psi}}^\star_{GMM}$ approaches infinity. This result provides a theoretical justification to the simulation results in Bramoullé et al. (2009) showing that an increase in the graph's density decreases the precision of the peer and contextual effects estimators. Under this framework, denser networks provide fewer opportunities to form moment conditions (see Figure 3 for an example). Theorem 2 exposes a relationship between the network parameters' convergence rate, precision, and network sparsity. To the best of my knowledge, this result is new in the literature of econometrics of networks.

5.1 Covariance Matrix Estimation

The estimation of the asymptotic variance-covariance matrix $\mathbf{\Sigma}_N^\star$ follows the network HAC estimator proposed by Kojevnikov et al. (2020). This class of covariance matrix estimators are formed by taking weighted averages of the network autocovariance terms with weights that are zero if the distance between two nodes is greater than D (see Assumption 2). In principle, as in Kojevnikov et al. (2020), the constant D can be a function of the sample size. The covariance matrix estimator is given by

$$\widehat{\mathbf{\Omega}}_N = \sum_{d\geq 0} K(d/D_n)\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{P}_n(i,d)} \mathbf{z}_{N,i}\widehat{\varepsilon}_{N,i}\widehat{\varepsilon}_{N,j}\mathbf{z}_{N,j}^\top .$$

where $\widehat{\varepsilon}_{n,i} = y_{n,i} - \mathbf{d}_{n,i}^\top \widehat{\boldsymbol{\psi}}^\star_{GMM}$, and $K(d)$ represents the wight associated with the size of the correlation between observed and unobserved characteristics of individuals i and j who are at distance d. The kernel function is such that $K(0) = 1$ and $K(x) = 0$ for $x > 1$. Kojevnikov et al. (2020) suggests a set of possible kernel functions that guarantee the expected properties for the covariance matrix estimator. The choice of the bandwidth D_n, and whether or not it depends on the sample size, is directly connected to the choices of K_c and K_d, and whether or not they depend on the sample size. From this estimator, it follows that the efficient variance-covariance matrix $\mathbf{\Sigma}_n^\star$ can be estimated by

$$\widehat{\mathbf{\Sigma}}_n^\star = \left[n^{-1}\mathbf{Z}_n^\top \mathbf{D}_n \widehat{\mathbf{\Omega}}_n^{-1} n^{-1}\mathbf{Z}_n^\top \mathbf{D}_n\right]^{-1}. \tag{13}$$

Standard Errors Calculation

The standard errors for the coefficient of interest can be computed by taking the squared-root of the main diagonal elements of the matrix $\widehat{\boldsymbol{\Sigma}}_n^{\star}$ after dividing them by n. The estimation of the efficient weighting matrix in (13) requires a consistent estimator of $\boldsymbol{\psi}$. A standard two-step GMM approach is used, where the one-step estimator for $\boldsymbol{\psi}$ is two stage least squares.

With this method, it is possible to efficiently estimate the parameters of the MLiM model while at the same time adjusting for multilayer network dependence. These estimation results show that in general, when estimating peer and contextual effects, the researcher should expect larger variances when the network density or sparsity increases. These results are novel and relevant for empirical work concerning network effects.

6 Monte Carlo Simulations

This section presents the simulation experiments designed to test the finite sample properties of the GMM estimator proposed in Theorem 2 and its robustness to endogenous multilayer network formation. I use a data generating process that mirrors the endogenous multilayer formation rule described in Appendix A. In particular, the number of regressors in equation (1) are defined as $Q = 1$. I assume that the same observed characteristic x_i affects both the outcome in (1) and the network formation rule in (A-19). The number of exogenous clusters are set to be $K = 10$. I randomly separate nodes into the ten clusters. For each cluster $k \in \{1, \ldots, 10\}$, the unobserved random vector $\mathbf{g}_k$ of size n_k is drawn from a multivariate normal with mean μ_k, variance 1, and inter-cluster correlation of 0.9 for all clusters k. I draw the 10×1 vector of cluster means $\boldsymbol{\mu}$ from a multivariate normal with mean 5, variance 2, and correlation 0.6. The idea of randomly drawing the cluster means is to generate some variation in how close the clusters are to each other. Finally, the vector of observed characteristics for cluster k is generated as $\mathbf{x}_k = \mathbf{g}_k + \boldsymbol{\epsilon}_{1,k}$, where $\boldsymbol{\epsilon}_{1,k}$ follows a multivariate normal standard normal with correlation 0.6 for all clusters.

I assume that the total number of layers is $M = 3$, and I form the network following the rule described in equation (A-19). Based on the resulting network, the outcome vector $\mathbf{y}$ is constructed following equation (3) with parameters $[\beta_1^0, \beta_2^0, \beta_3^0, \delta_1^0, \delta_2^0, \delta_3^0, \gamma^0, \alpha^0] = [0.1, 0.2, 0.3, 1, 2, 3, 2, 1]$, where $\varepsilon_i = g_i + \epsilon_{2,i}$, and $\epsilon_{2,i}$ follows a standard normal distribution. The fact that both ε_i and x_i depend on g_i induces endogeneity in the model. Given that the cluster membership within the data generating process (DGP) are known, and the ten clusters created are independent to each other, it is possible to exactly calculate the values of K_c and K_d that make equations (6) and (7) hold. Therefore, I can form the correctly specified matrices of moment conditions given by $\mathcal{W}_{m,\beta}$ and $\mathcal{W}_{m,\delta}$ for $m = \{1, 2, 3\}$. Note that the values of K_c and K_d depend on the realization of the random variables presented before, and therefore change from iteration to iteration. There are cases where given K_c and K_d, there are not enough identifying variation to estimate the parameters of interest. To solve this issue, the DGP is simulated until I reached a total of 1,000 data sets.

Based on the 1,000 data sets $\left\{y_{n;i}, x_{n;i}, \{w_{n,1;i,j}\}_{j=1,j\neq i}^{n}, \{w_{n,2;i,j}\}_{j=1,j\neq i}^{n}, \{w_{n,3;i,j}\}_{j=1,j\neq i}^{n}\right\}_{i=1}^{n}$ with $n \in \{50, 100, 200\}$, the parameters of interest $[\beta_1^0, \beta_2^0, \beta_3^0, \delta_1^0, \delta_2^0, \delta_3^0, \gamma^0, \alpha^0]$ are estimated by using the proposed GMM estimator. Given that $Q = 1$, in this case, the system is just identified, and concerns about efficiency can be disregarded. Figure 4 presents the results for the Monte Carlo simulation for the parameters $[\beta_2^0, \delta_2^0, \gamma^0]$. The result for the other parameters can be found in the Online Appendix D. The first row in Figure 4 displays the performance of the GMM where the box plots are shown with whiskers displaying the 5% and 95% empirical Monte Carlo quantiles. Across the board, for all parameters, the proposed GMM estimator performs well in terms of bias and sampling variability for sample sizes as small as 50 nodes. As expected, the estimation variability decreases when increasing the sample size. Similarly, the second row of Figure 4 displays the corresponding Q-Q plots for the GMM based on the standardized version of the Monte Carlo replications for sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black). The blue dashed line depicts the 45 degree line. This plot shows that the asymptotic normal approximation in Theorem 2 works well even with small samples. The approximation improves when the sample size increases. These results confirm that the proposed GMM estimator is robust to multilayer network formation process, under the assumption that individuals' dependence decreases with their distance in the multilayer network space.

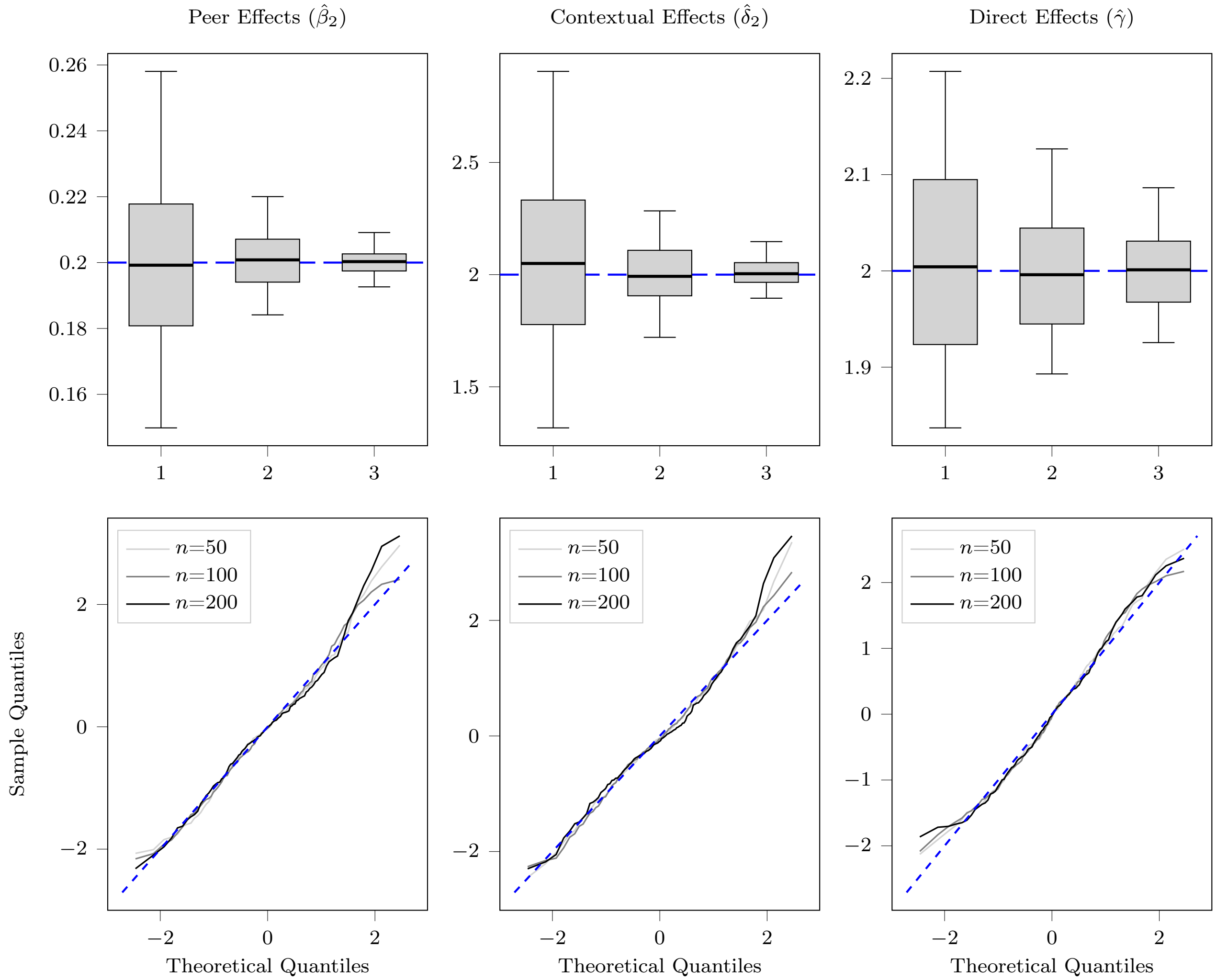
7 Application to Publication Outcomes in Economics

To illustrate the behavior of the efficient GMM estimator presented in Theorem 2, data of 1,628 peer-reviewed articles published between 2000-2006 in the top-four general-interest journals in Economics is used. The availability of online scientific research repositories has resulted in a stable source of data to uncover scholars' professional connections. Additionally, when linked with scholars' biographical public information, other type of professional connections such as Ph.D. alumni, colleagues and advisorship can be uncovered (Colussi, 2018). A major challenge when performing causal inference analysis with these data is that the professional networks that scholars form are inherently correlated with publications outcomes such as citation counts. Therefore, this is an ideal scenario where the proposed method can be used to uncover network effects and attach them causal interpretation using observational data.

7.1 Data on Publication Outcomes

The data consists of all peer-reviewed articles published in the top 4 general-interest journals in Economics, namely the *American Economics Review* (AER), *Econometrica* (ECA), the *Journal of Political Economy* (JPE), and the *Quarterly Journal of Economics* (QJE) between 2000 and 2006. This information was web scrapped from three different sources: Ideas RePEc, Scopus, and the journal websites themselves. This allows to correctly identify editors' names and tenure, as well as research articles'

Figure 4: Box Plots and Q-Q plots of the GMM Estimator



Note: Box plots in the first row depict the Monte Carlo performance of the proposed GMM estimator. The boxplots are based on 1,000 for sample sizes $n \in \{50, 100, 200\}$. The whiskers display the 5% and 95% empirical quantiles. Q-Q plots in the second row are based on the standardized sample of 1,000 Monte Carlo replications of the proposed GMM estimator of the parameters in (1) and sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black). The blue dashed line shows the 45 degree line.

titles, page numbers, total number of bibliographic references, complete authors' names and affiliation at the time of publication, authors-provided Journal of Economic Literature (JEL) codes (AER, QJE), and keywords (ECA). After excluding editorial reports, conference announcements and proceedings, corrigendums, comments, replies, special issues, and Nobel prize lectures reprints, a total of 1,628 articles are used. This roughly coincides with the 1,657 articles compiled by Card and DellaVigna (2013), and the 1,620 papers identified by Colussi (2018) for the same journals and time period. The total number of citations 8 years post publication for each article was then extracted from Scopus and completed with that of Colussi (2018). Similarly, having identified 1,988 unique authors and 42 unique editors (37 of which also published papers in these journals in this time period), information regarding their gender, research interests (coded as a JEL code), education, and employment history was obtained from online profiles and online curricula vitae by means of web scrapping and text mining. This information was then merged with similar attributes already collected by Colussi (2018) for 1,882 of them.

Table 1 presents descriptive statistics. It shows that the number of peer-reviewed research articles published in these journals is relatively stable in this time period, and that the AER has the largest share of about 40% of the total number of articles for each year. There was an average of 1.8 authors per paper in this time period, and the average number of pages per article has slightly increased from 25 in 2002 to 27 in 2006. All these findings coincide overall with the facts presented by Card and DellaVigna (2013) in a larger time period including ours. The total number of citations up to 8 years post publication allows direct comparisons of the impact of papers published at different points in time. The average citation per article is relatively constant with an average of 60 citations per paper. The QJE had a higher-than-average citation per article relatively to the other journals in 2002 and 2003. Authors' gender allows to identified articles written by same-gender authors (males or females only) as well as different-gender authors (males and females). Different-gender articles represent about 14% of the total number of articles and this proportion does not show any particular trend in this time period.

7.2 Multilayer Network Data

The multilayer network is composed of four different types of connections among authors and editors (scholars hereafter), and it can be constructed using their co-authorship information, research interests, education, and employment history. In particular, the *Co-author* layer is constructed by defining connections (edges) between scholars l and k if they co-authored a paper together. The *Alumni* layer is constructed by creating edges between scholars l and k if both obtained their Ph.D. from the same institution during the same time window. The *Ph.D. Advisor* connection is constructed by creating a link between two scholars if one of them held the rank of Assistant, Associate, or Professor at the same university and in the same year in which the other obtained his or her Ph.D, and they share least one research field. Finally, two scholars are said to be *Colleagues* if they worked in the same institution in the same time period. Details of how these two types of academic ties among scholars are constructed

Table 1: Descriptive Statistics for Research Articles

Variable	2000	2001	2002	2003	2004	2005	2006
Number of Authors	376	410	486	411	395	410	411
AER	136	166	189	168	140	169	172
ECA	93	111	153	107	107	111	108
JPE	87	79	93	80	98	76	71
QJE	74	80	73	80	80	82	84
Number of Articles	221	238	270	232	226	225	216
AER	79	88	92	90	79	89	88
ECA	51	64	90	59	61	55	51
JPE	49	44	48	43	46	41	37
QJE	42	42	40	40	40	40	40
Number of Citations	10,856	12,088	14,161	14,079	14,564	14,861	13,103
AER	4,140	4,833	4,476	4,586	5,354	5,606	5,289
ECA	2,530	2,773	3,346	3,379	2,811	2,724	2,914
JPE	1,692	1,446	2,240	1,954	2,787	3,420	1,852
QJE	2,494	3,036	4,099	4,160	3,612	3,111	3,048
Number of Editors	16	20	23	16	18	24	14
AER	3	2	7	6	4	8	9
ECA	4	9	5	3	8	2	3
JPE	4	2	7	4	3	7	1
QJE	5	7	4	3	3	7	1
Number of Pages	5,486	5,984	6,870	6,327	6,356	6,529	5,979
AER	1,409	1,537	1,557	1,636	1,589	1,702	1,807
ECA	1,297	1,566	2,341	1,735	1,772	1,877	1,559
JPE	1,337	1,379	1,424	1,396	1,429	1,397	1,131
QJE	1,443	1,502	1,548	1,560	1,566	1,553	1,482
Number of References	7,042	7,425	8,426	7,633	7,645	8,115	7,141
AER	2,420	2,783	2,729	2,896	2,576	3,032	2,888
ECA	1,388	1,695	2,598	1,772	2,013	1,886	1,467
JPE	1,285	1,311	1,452	1,391	1,503	1,445	1,166
QJE	1,949	1,636	1,647	1,574	1,553	1,752	1,620
Different Gender	24	34	37	34	29	27	41
AER	7	15	18	16	9	15	21
ECA	7	6	8	5	5	2	9
JPE	4	6	5	5	7	0	3
QJE	6	7	6	8	8	10	8

Note: Descriptive statistic for articles published in the *American Economics Review* (AER), *Econometrica* (ECA), the *Journal of Political Economy* (JPE), and the *Quarterly Journal of Economics* (QJE) between 2000 and 2006. 'Number of Authors' counts the total number of unique authors that participated in the writing of the articles presented in the variable 'Number of Papers.' 'Number of Citations' refers to the total citations up to 8 years post publication. 'Number of Editors' shows the total number editors in these journals with at least 1 year tenure in the time period. 'Number of References' counts the total number of items that each paper cites in its bibliographic references section, and 'Different Gender' counts the total number of articles written by co-authors of different gender.

can be found in the Online Appendix D.

Since articles' authors are observed, these four types of professional ties can then be collapsed at the article level instead, i.e., articles i and j are connected in networks $\mathbf{W}_{n,m}$ if at least one of the authors of article i shares a m-type connection with at least one of the authors of article j respectively. These networks are formed in a cumulative way, e.g., the 459 articles (nodes) in 2001 include the 221 articles published in 2000 and so on till all 1,628 articles are accounted for in 2006.

7.3 Estimation of Network Effects in Publication Outcomes

This section aims at quantifying the potential existence of human capital externalities (peer effects) among scholars publishing in the 4 top general-interest journals described above. Specifically, the running hypothesis is that if a paper's authors are connected with other scholars who produce good quality articles (measured in terms of citations, see, e.g., Card et al. 2020), the quality of their own article will increase because of the existence of strategic complementarities, see., e.g., Boucher and Fortin (2016). The previously defined professional connections of co-authorship ($m = 1$), alumni ($m = 2$), advisorship ($m = 3$), and colleagues ($m = 4$) are taken into account in the following specific case of the estimating equation (1):

$$y_{i,r,t} = \alpha + \sum_{m=1}^{3} \sum_{j \neq i} w_{m,i,j,t} y_{j,r,t} \beta_m + \sum_{m=1}^{3} \sum_{j \neq i} w_{m,i,j,t} \widetilde{\mathbf{x}}_{j,r,t}^{\top} \boldsymbol{\delta}_m + \mathbf{x}_{i,r,t}^{\top} \boldsymbol{\gamma} + \lambda_r + \lambda_t + \lambda_0 + \varepsilon_{i,r,t}, \qquad (14)$$

where $y_{i,r,t}$ represents the natural logarithm of the total number of citations up to 8 years post publication of article i, in journal r, at time t. The scalar $w_{m,i,j,t}$ represents the (i, j) entry of the $\mathbf{W}_m$ adjacency matrix for the *co-authorship*, *alumni*, *advisorship* and *colleagues* networks at time t. The controls in $\mathbf{x}_{i,r,t}$ include dummies for whether current or previous editors of journal r at time t are authors of article i (`Editor`), and for whether all the authors of article i have different gender (`Different Gender`). The latter is coded as zero for single-authored publications. Other articles characteristics are also included such as the total number of pages (`Number of Pages`), total number of authors (`Number of Authors`), total number of bibliographic references (`Number of References`), and another dummy that equals one for isolated articles in the three different networks (`Isolated`). Contextual effects are only calculated for the `Editor` and `Different Gender` covariates, i.e., $\widetilde{\mathbf{x}}_{j,r,t}$. Model (14) is estimated in a rolling-regression setting for $t = 2002$, 2003, 2004, 2005, and 2006, i.e., the estimating sample each year includes those from previous years. Results for the year 2000 and 2001 are not included because they suffer degrees-of-freedom problems given specification (14). Given that the estimator in this paper is designed for cross-sectional data, I pull all the years together and add year fixed effects (λ_t). Estimating equation (14) also includes journal (λ_r) fixed effects. As mentioned before, the scalar structural error $\varepsilon_{i,r,t}$ is such that $\mathbb{E}(\boldsymbol{\varepsilon}|\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_4) \neq 0$ because of the potential endogeneity of the professional networks included in the estimation. A potential reason to be concerned about network endogeneity is that authors producing high quality papers may be connected with each other just because they are similar in their labels of skills, i.e., peer effects can be confounded with unobserved heterogeneity or homophily.

Therefore, to be able to provide a causal interpretation for the peer effect parameters in equation (14), one needs to impose restrictions on the dependence patters on the population joint distribution $\mathcal{F}$. For this empirical application, the constants K_c and K_d are chosen to be $K_c = 1$ and $K_d = 3$. This choice of parameters implies that individuals who are connected by paths with one edge type change or at least three edges are assumed to be uncorrelated. Therefore, changes in their exogenous characteristics can be used as exogenous variation to identify the parameters of interest. In particular, I use the `Different Gender`, `Number of Pages`, `Number of Authors` and `Number of References` as the exogenous characteristics to form the moment conditions in equations (6) and (7). Table 2 presents the results for the Efficient GMM estimator characterized in Theorem 2. For comparison, the results for the first stage GMM estimator and the OLS estimator are presented in Tables 3 and 4 in the Online Appendix D. The analysis is conducted by changing the constants K_c and K_d to be $K_c = 2$ and $K_d = 4$. The magnitude and direction of the estimated parameters in Table 2 are robust to changes in these constants. However, as suggested by the identification and asymptotic theory, the standard errors of the estimated coefficients increase. The reason is that the amount of information that can be used to identify and estimate the parameters of interest is decreasing in K_c and K_d. When trying to perform the analysis for values of $K_c = 3$ and $K_d = 5$, all the standard errors greatly surpass the value of the estimated parameters, this presents empirical evidence of the remarks in Theorem 2 arguing that when the probability of finding moment conditions approaches zero, the parameters' variance-covariance matrix could grow arbitrarily large.

Regarding the estimators' behavior, Tables 2 and 3 confirms the results predicted by the estimation theory. Across the board, the estimated standard errors of the Efficient GMM estimator are lower than those of the first stage GMM. The coefficients estimated by the first stage GMM and the Efficient GMM are relatively consistent, which shows empirical evidence of the consistency results in Theorem 2. Additionally, the OLS estimated peer and contextual effects parameters in Table 4 differ from those estimated by the proposed GMM method. This result is expected. As argued before, the layers included in the estimation are likely to be formed endogenously.

In terms of empirical findings, Table 2 provides different meaningful results. First, building upon Ductor et al. (2014), peer effects are found to be positive and statistically significant for articles' quality coming from the *Co-authors* network for all years. This positive result can potentially reflect human capital spillovers that could act through a variety of mechanisms, e.g., scholars provide feedback on each other's work, serve as referees, and work directly together. These instances of collaboration are often paramount to the extension of existing research agendas and can bring to light novel ideas. However, the peer effects estimators from the other professional networks do not significantly affect the publication outcomes. This result is new to the literature of scholars' research productivity. It emphasizes the importance of a network that guarantees a direct channel of communication between authors instead of other professional networks that may generate fewer interpersonal interactions.

The `Editor` contextual effects are, in general, insignificant for all the networks included in the regression. For each article, this variable counts the number of authors who are or have been editors of at least one of the 4 journals till the relevant year. Then, the contextual effect for article i measures the influence on quality of the average number of editors that have written papers that are connected to i in one of the networks of interest. The intuition here is that editors may have superior information of the publication process, they have greater recognition, and therefore they can potentially produce highly cited papers. However, after controlling for peer effects and other article characteristics, the editor contextual effects are not significantly different from zero across the two networks. This is an interesting result because there has been evidence suggesting that authors connected to editors are more likely to publish their work (see, e.g., Laband and Piette 1994), however, when it comes to citations, the results suggest that articles connected to an editor do not benefit from such a connection. The same is true for potential gender contextual effects, except for the advisor network between the years 2003 and 2005. Having advisors that work in projects with gender diverse groups, seems to have a positive impact on their students publications' outcomes.

Similarly, a larger number of bibliographic references is associated with greater impact in terms of citations for estimators. Finally, the direct effects of `Different Gender` on the articles' quality are all positive and significant for most years. Articles written by authors of different gender have a significantly higher number of citations than articles written by same-gender authors. On average, different-gender articles will have somewhere between 24% and 26% more citations than same-gender articles holding everything else constant. This finding is robust across all estimators and presents evidence of an improvement in the quality of the papers when the research teams are gender diverse after controlling for peer effects, editorial participation, article characteristics, and a complete set of fixed effects. This finding is in line with recent research investigating differences in publication outcomes by gender, see, e.g., Card et al. 2020.

8 Conclusion

This paper provides a novel approach to show how multilayer network data structures can be used to identify and consistently estimate network effects. The proposed method applies to an extension of the linear-in-means model which relaxes the assumption that only one type of network can generate peer and contextual effects. This paper's results provide a tool to identify multilayer network effects in settings with endogenous network formation and network dependence between observed and unobserved individual characteristics. I show that it is possible to identify the model by imposing mild conditions on the dependence structure of the multilayer network space. In particular, I impose $\psi-$weak dependence to model the correlation structure. This work provides a simple linear GMM estimator and characterize its

Table 2: Efficient GMM Estimation Results for Social and Direct Effects

	2002	2003	2004	2005	2006
Peer Effects ($\{\widehat{\beta}\}_{m=1}^{3}$)					
`Co-authors`	0.425***	0.453***	0.574***	0.534***	0.523**
	(0.117)	(0.107)	(0.111)	(0.103)	(0.271)
`Alumni`	0.125	0.128	0.195	0.148	0.369
	(0.237)	(0.201)	(0.237)	(0.164)	(0.381)
`Advisor`	-0.133	0.192	-0.655**	-0.369	-0.105
	(0.229)	(0.253)	(0.316)	(0.227)	(0.360)
`Colleagues`	0.514	-0.233*	-0.054	0.051	0.092
	(0.406)	(0.111)	(0.096)	(0.096)	(0.155)
Contextual Effects ($\{\widehat{\boldsymbol{\delta}}\}_{m=1}^{3}$)					
`Co-authors: Editor in Charge`	0.059	-0.281	-0.326	-0.465	0.879
	(0.657)	(0.589)	(0.655)	(0.542)	(0.952)
`Alumni: Editor in Charge`	-0.031	0.055	0.019	0.051	0.417
	(0.201)	(0.150)	(0.136)	(0.125)	(0.312)
`Advisor: Editor in Charge`	0.678	0.257	1.585**	0.975	0.431
	(0.662)	(0.710)	(0.726)	(0.653)	(0.774)
`Colleagues: Editor in Charge`	0.348	0.006	-0.225	-0.271	-0.208
	(0.258)	(0.179)	(0.169)	(0.182)	(0.394)
`Co-authors: Different Gender`	-0.462	-0.541	-1.548	-1.089	3.001
	(1.508)	(1.117)	(1.496)	(0.827)	(1.776)
`Alumni: Different Gender`	-0.439	-0.207	-0.329	0.039	-1.041
	(0.401)	(0.363)	(0.372)	(0.288)	(0.866)
`Advisor: Different Gender`	0.238	4.096***	6.062***	4.224***	1.268
	(2.735)	(1.566)	(1.740)	(1.443)	(1.764)
`Colleagues: Different Gender`	0.435	0.169	-0.073	-0.382	-2.710
	(0.476)	(0.543)	(0.554)	(0.712)	(1.636)
Contextual Effects ($\widehat{\gamma}$)					
`Editor in Charge`	-0.099	-0.081	-0.042	-0.029	0.011
	(0.125)	(0.111)	(0.133)	(0.118)	(0.138)
`Different Gender`	0.247**	0.238**	0.182**	0.127	0.080*
	(0.115)	(0.097)	(0.089)	(0.078)	(0.092)
`Number of Pages`	0.023***	0.022***	0.019***	0.017***	0.016***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
`Number of Authors`	0.067	0.088	0.046	0.072*	0.069**
	(0.053)	(0.045)	(0.043)	(0.038)	(0.035)
`Number of References`	0.009***	0.009***	0.007***	0.009***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
`Co-authors: Isolated`	1.427***	1.477***	1.789***	1.642***	1.661*
	(0.414)	(0.388)	(0.398)	(0.372)	(1.001)
n	729	961	1187	1412	1628

Note: Standard errors are in parenthesis and are calculated using the network HAC estimator of the covariance matrix in equation (5.1) where the function K is the Parzen kernel and the bandwidth $D_n = 1.8 \times [\log(\text{ avg.deg } \vee (1.05))]^{-1} \times \log n$. Stars follow the key: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$, where p stands for p-values. R^2 are calculated as the squared of the sample correlation coefficients between the observed outcomes and their fitted values. All specifications include indicator variables for Journal and Year. The indicator for isolated nodes in the Alumni, Advisor and Colleagues networks are also included but are not statistically significant.

limiting distribution. The asymptotic covariance matrix accounts for the multilayer network dependence and incorporates the possibility that too dense or too sparse networks could provide weak identifying information, increasing the estimator's variance. These results regarding the asymptotic covariance matrix allow constructing correct standard errors for inference.

The proposed method will be helpful to empirical researchers aiming to estimate network effects using observational data. This framework can handle both monolayer and multilayer data structures. However, it is more useful when the analyst observes at least two layers. The reason is that the proposed method is designed to use all the additional information provided by including additional types of connections. In addition, this method is best suited for studies where the researcher has access to a credible exogenous shock. The shocks do not need to be strictly exogenous but rather exogenous from the perspective of other individuals far away in the multilayer network space. This characteristic of the method resembles the partial population identification ideas (Moffitt, 2001).

To showcase the applicability of the proposed method, this work presents an empirical application where the tools of web scraping and text mining are used to construct a data set consisting of all peer-reviewed research articles published in 4 top general-interest economics journals between 2000 and 2006. Using publicly available information on where the authors of these publications obtained their Ph.D. degree from, along with their working history, I construct the *Alumni*, *Ph.D. Advisor*, and *Colleagues* layers, in addition to the *Co-authorship* layer. Results show the existence of positive peer effects in terms of citations among peer-reviewed research articles connected through co-authorship connections of their authors as well as significant positive effects of research teams that are gender diverse on the quality of a paper measured in terms of citation outcomes. I do not find evidence of network effects from any of the other layers. I interpret this result as emphasizing the importance of a network that guarantees a direct communication channel between authors instead of other professional networks that may generate fewer interpersonal interactions. I find evidence of bias in the OLS estimator and increasing standard errors corresponding to increases in the values of K_c and K_d (as predicted by the asymptotic theory).

While this paper provides new results in the econometrics of networks literature, it also opens the possibility for new research. Future work could include the explicit characterization of the relationship between the values of the matrices $\mathcal{K}_{N,\lambda,i}$ and $\mathcal{K}_{N,i}$ and the denseness or sparsity of the multiyear network by potentially assuming a general multilayer network formation process. Another research avenue can be to investigate how to optimally choose the hyperparameters controlling the network dependence or the weights used in the moment conditions.

9 Proofs of Main Results

Proof of Proposition 1. Choose f such that it selects an arbitrary position q from the vector $\mathbf{r}_{N,i}$, i.e., $f(\mathbf{r}_{N,i}) = V_q\mathbf{r}_{N,i} = x_{N,i,q}$ where $x_{N,i,q}$ denotes the qth regressor in $\mathbf{x}_{N,i}$. Similarly, choose g to select the $Q+1$th position of $\mathbf{r}_{N,i} = V_{Q+1}\mathbf{r}_{N,i} = \varepsilon_{N,i}$. Note that $\|f\|_\infty = |x^q_{N,i}| < \infty$, $\|g\|_\infty = |\varepsilon_{N,i}| < \infty$,

$\text{Lip}(f) = \text{Lip}(g) = 1$, so that from Assumption 2 (a) and (b):

$$|\text{Cov}\,(x_{N,i,q}, \varepsilon_{N,i})| \leq C\,(|x_{N,i,q}| + 1)\,(|\varepsilon_{N,i}| + 1)\,\theta_{N,s}.$$

By Assumption 2 part *(iii)*, it follows that if $d_N^{\mathcal{M}}(i,j) \geq D$, then $\theta_{N,s} = 0$. Set $K_d, K_c \in \mathbb{N}_+$ such that $K_d > K_c + 1$, let $D = K_d$ and $\tau = \mathbb{1}\{d_N^c(i,j) \geq K_d\}K_c(K_d - K_c - 1)$. Note that if $K_d \leq K_c + 1$ the condition in equation (7) does not provide any additional information. The reason is that by definition, the shortest path $d_N^*(i,j)$ and the number of edge type changes $c_N^*(i,j)$ are such that $d_N^*(i,j) \geq c_N^*(i,j)$. Let $K_d \leq K_c + 1$, then $c_N^*(i,j) \geq K_c$ directly implies $d_N^*(i,j) \geq K_d$. Thus, the inequality $c_N^*(i,j) \geq K_c$ does not provide additional information. If $c_{i,j}^* \geq K_c$ and $d_N^c(i,j) \geq K_d$, then $d_N^{\mathcal{M}}(i,j) \geq d^*(i,j) + K_c^0(D - K_c - 1) > D$ implying $\theta_{n,s}$=0. Similarly, if $c^*(i,j) < K_c$ and $d^*(i,j) \geq K_d$, then $d_N^{\mathcal{M}}(i,j) = D + \tau c_N^*(i,j) > D$, which also implies $\theta_{n,s} = 0$. By Assumption 3, $\text{Cov}(x_{N,i,q}, \varepsilon_{N,i}) = \mathbb{E}(x_{N,i,q}, \varepsilon_{N,i})$, so that equations (6) and (6) hold for $x_{N,i,q}$. Because q was chosen arbitrarily, the result holds for all q in $\mathbf{x}_i$ completing the proof. ∎

Proof of Theorem 1. First note that Assumption 1 guarantees that the solution for model (1) exists. Fix K_c and K_d with $K_d > K_c + 1$, and let Assumption 2 and 3 holds such that Proposition 1 follows for any realization of $\mathcal{M}_N \in \mathscr{M}$. Assumption 4 part *(ii)* guarantees that there is at least one individual for which it is possible to form a moment condition. Combining the results of Proposition 1 with the law of iterated expectations, it follows that $\mathbb{E}[\mathbf{m}(\boldsymbol{\psi}^0)] = 0$. Choose an arbitrary vector of parameters $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ such that $\mathbb{E}[\mathbf{m}(\boldsymbol{\psi}^0)] = 0$. Notice that $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}(y_{N,i} - \mathbf{d}_{N,i}^\top\boldsymbol{\psi})] = 0$ implies $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top](\boldsymbol{\psi}^0 - \boldsymbol{\psi}) + \mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\varepsilon_{N,i}] = 0$, and $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]\left(\boldsymbol{\psi}^0 - \boldsymbol{\psi}\right) = 0$. Under the Assumption 5, it follows that $\mathbb{E}[\mathbf{m}(\boldsymbol{\psi})] = 0$ if and only if $\boldsymbol{\psi}^0 = \boldsymbol{\psi}$. ∎

Proof of Theorem 2. The GMM estimator from Section 5 in the main text can be written as

$$\widehat{\boldsymbol{\psi}}_{GMM} = \boldsymbol{\psi} + (n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\left(\mathbf{A}_n^\top\mathbf{A}_n\right)n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n)^{-1}n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n(\mathbf{A}_n^\top\mathbf{A}_n)n^{-1}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n \tag{15}$$

By construction, the matrix $\mathbf{A}_n$ is assumed to converge to $\mathbf{A}_N$, so that $(\mathbf{A}_n^\top\mathbf{A}_n) \rightarrow (\mathbf{A}_N^\top\mathbf{A}_N)$ as $n \rightarrow \infty$, which is assumed to be finite and full rank. From Lemma 6, $n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n$ converges to the population quantity $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]$, which is finite given Assumption 5. Finally, Lemma 7 shows that $n^{-1}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n(\boldsymbol{\psi})$ converges to $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})] = 0$. It follows that $\widehat{\boldsymbol{\psi}}_{GMM} = \boldsymbol{\psi} + o_p(1)$. For asymptotic normality, note that, from equation 15

$$\sqrt{n}(\widehat{\boldsymbol{\psi}}_{GMM} - \boldsymbol{\psi}) = (n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\left(\mathbf{A}_n^\top\mathbf{A}_n\right)n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n)^{-1}n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n(\mathbf{A}_n^\top\mathbf{A}_n)n^{-1/2}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n.$$

Let $\mathbf{Q}_{zx} = \mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]$. From Lemmas 6 and 8 it follows that

$$\sqrt{n}\left(\widehat{\psi}_{GMM} - \psi\right) \xrightarrow{d} \left[\mathbf{Q}_{zx}^\top\left(\mathbf{A}^\top\mathbf{A}\right)\mathbf{Q}_{zx}\right]^{-1}\mathbf{Q}_{zx}^\top\left(\mathbf{A}^\top\mathbf{A}\right)N(\mathbf{0}, \boldsymbol{\Omega}_N),$$

The efficient weighing matrix is given by $\mathbf{A}_N = \mathbf{\Omega}_N^{-1/2}$ so that $\mathbf{A}_N^\top \mathbf{A}_N = \mathbf{\Omega}_N^{-1}$. With that choice of weighting matrix, the asymptotic variance-covariance matrix is given by

$$\mathbf{\Sigma}_N^\star = [\mathbf{Q}_{zx}^\top \mathbf{\Omega}_N^{-1} \mathbf{Q}_{zx}]^{-1}$$

By the remarks presented in section 4 about conditional expectation interpretation of $\mathbf{Q}_{zx}$, it follows that $\mathbf{\Sigma}_N^\star$ can also be written as

$$\left[\left(\sum_{i\in\mathcal{I}_N} \mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top \mid \mathcal{H}_{N,\lambda,i}^*, \mathcal{H}_{N,i}^* \neq \mathbf{O}_{R\times R}]\mathcal{K}_{N,i}\right)^\top \mathbf{\Omega}_N^{-1} \left(\sum_{i\in\mathcal{I}_N} \mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top \mid \mathcal{H}_{N,\lambda,i}^*, \mathcal{H}_{N,i}^* \neq \mathbf{O}_{R\times R}]\mathcal{K}_{N,i}\right)\right]^{-1}, \tag{16}$$

Equation (16) shows that the efficient asymptotic variance-covariance matrix of the coefficient vector $\widehat{\psi}_{GMM}$ can grow arbitrarily large if the matrices $\mathcal{K}_{N,\lambda,i}^*$ and $\mathcal{K}_{N,i}^*$ (containing the upper right sub-matrices related with the values of $\kappa_{N,m,\lambda,i}$ and N, m, i) are too close to the zero matrix. I interpret this result as week identification of the peer and contextual effects parameters when the probabilities of finding moment conditions are low. Those probabilities are linked with the density/sparsity of the population network. The efficient GMM estimator is given by $\widehat{\boldsymbol{\psi}}_{GMM}^\star = (\mathbf{D}_n^\top \mathbf{Z}_n \mathbf{\Omega}_n^{-1} \mathbf{Z}_n^\top \mathbf{D}_n)^{-1}(\mathbf{D}_n^\top \mathbf{Z}_n \mathbf{\Omega}_n^{-1} \mathbf{Z}_n^\top \mathbf{y}_n)$. The previous arguments imply that $\sqrt{n}(\widehat{\boldsymbol{\psi}}_{GMM}^\star - \boldsymbol{\psi}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_N^\star)$. Note that the consistency argument also applies to $\widehat{\boldsymbol{\psi}}_{GMM}^\star$ given that by Lemma 9, $\mathbf{\Omega}_n \to \mathbf{\Omega}_N < \infty$, as $n \to \infty$. ■

References

Amini H, Lelarge M. 2015. The diameter of weighted random graphs. *The Annals of Applied Probability* **25**: 1686–1727.

Auerbach E. 2016. Identification and estimation of models with endogenous network formation. Unpublished Manuscript.

Balasubramanian V, Estrada J, Ho ATY, Huynh KP, Jacho-Chávez DT. 2022. Searching for shortest paths in multilayered networks. going beyond the local neighbourhood. Unpublished Manuscript.

Blume LE, Brock WA, Durlauf SN, Jayaraman R. 2015. Linear social interactions models. *Journal of Political Economy* **123**: 444–496.

Boccaletti S, Bianconi G, Criado R, del Genio CI, Gómez-Gardeñes J, Romance M, Sendiña-Nadal I, Wang Z, Zanin M. 2014. The structure and dynamics of multilayer networks. *Physics Reports* **544**: 1–122.

Boucher V, Fortin B. 2016. Some challenges in the empirics of the effects of networks. *The Oxford handbook of the economics of networks* : 277–302.

Bramoullé Y, Djebbari H, Fortin B. 2009. Identification of peer effects through social networks. *Journal of Econometrics* **150**: 41–55.

Bramoullé Y, Djebbari H, Fortin B. 2020. Peer effects in networks: a survey. Unpublished Manuscript.

Bursztyn L, Ederer F, Ferman B, Yuchtman N. 2014. Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica* **82**: 1273–1301.

Burt RS. 2004. Structural holes and good ideas. *American journal of sociology* **110**: 349–399.

Calvó-Armengol A, Patacchini E, Zenou Y. 2009. Peer effects and social networks in education. *The review of economic studies* **76**: 1239–1267.

Card D, DellaVigna S. 2013. Nine facts about top journals in economics. *Journal of Economic Literature* **51**: 144–61.

Card D, DellaVigna S, Funk P, Iriberri N. 2020. Are Referees and Editors in Economics Gender Neutral? *The Quarterly Journal of Economics* **135**: 269–327.

Chan TJ, Estrada J, Huynh KP, Jacho-Chávez DT, Lam CT, Sánchez-Aragón L. 2022. On the estimation of social effects with observational network data and random assignment. Unpublished Manuscript.

Christakis N, Fowler J, Imbens GW, Kalyanaraman K. 2020. An empirical model for strategic network formation. In *The Econometric Analysis of Network Data*. Elsevier, 123–148.

Chung F, Lu L. 2001. The diameter of sparse random graphs. *Advances in Applied Mathematics* **26**: 257–279.

Colussi T. 2018. Social ties in academia: A friend is a treasure. *The Review of Economics and Statistics* **100**: 45–50.

Conley TG. 1999. Gmm estimation with cross sectional dependence. *Journal of econometrics* **92**: 1–45.

Conley TG, Udry CR. 2010. Learning about a new technology: Pineapple in ghana. *American economic review* **100**: 35–69.

De Giorgi G, Frederiksen A, Pistaferri L. 2020. Consumption network effects. *The Review of Economic Studies* **87**: 130–163.

De Giorgi G, Pellizzari M, Redaelli S. 2010. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics* **2**: 241–75.

de Paula Á. 2017. *Econometrics of Network Models*, volume 1 of *Econometric Society Monographs*, chapter 8. Cambridge University Press, 268–323.

De Paula Á, Richards-Shubik S, Tamer E. 2018. Identifying preferences in networks with bounded degree. *Econometrica* **86**: 263–288.

Doukhan P, Louhichi S. 1999. A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications* **84**: 313–342.

Ductor L, Fafchamps M, Goyal S, van der Leij MJ. 2014. Social networks and research output. *The Review of Economics and Statistics* **96**: 936–948.

Gerschgorin S. 1931. Uber die abgrenzung der eigenwerte einer matrix. *Izvestija Akademii Nauk SSSR, Serija Matematika* **7**: 749–754.

Goldsmith-Pinkham P, Imbens GW. 2013. Social networks and the identification of peer effects. *Journal of Business and Economic Statistics* **31**: 253–264.

Graham BS. 2015. Methods of identification in social networks. *Annual Review of Economics* **7**: 465–485.

Graham BS. 2017. An econometric model of network formation with degree heterogeneity. *Econometrica* **85**: 1033–1063.

Graham BS. 2020. Network data. *Handbook of Econometrics* **7A**.

Graham BS, Pelican A. 2020. Chapter 4 - testing for externalities in network formation using simulation. In Graham B, Áureo de Paula (eds.) *The Econometric Analysis of Network Data*. Academic Press, 63–82.

Jackson MO, Rogers BW, Zenou Y. 2017. The economic consequences of social-network structure. *Journal of Economic Literature* **55**: 49–95.

Jenish N, Prucha IR. 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of econometrics* **150**: 86–98.

Johnsson I, Moon HR. 2019. Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach. *The Review of Economics and Statistics* : 1–51.

Joshi S, Mahmud AS, Tzavellas H. 2020. “the tangled webs we weave": Multiplex formation. Unpublished Manuscript.

Kelejian HH, Prucha IR. 1998. 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**: 99–121.

Kelejian HH, Prucha IR. 2001. On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics* **104**: 219 – 257.

Kivela M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. 2014. Multilayer networks. *Journal of Complex Networks* **2**: 203–271.

Kojevnikov D, Marmer V, Song K. 2020. Limit theorems for network dependent random variables. *Journal of Econometrics* .

Kuersteiner GM. 2019. Limit theorems for data with network structure. *arXiv preprint arXiv:1908.02375* .

Kuersteiner GM, Prucha IR. 2020. Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica* **88**: 2109–2146.

Kuhn P, Kooreman P, Soetevent A, Kapteyn A. 2011. The effects of lottery prizes on winners and their neighbors: Evidence from the dutch postcode lottery. *American Economic Review* **101**: 2226–2247.

Laband DN, Piette MJ. 1994. Favoritism versus search for good papers: Empirical evidence regarding the behavior of journal editors. *Journal of Political Economy* **102**: 194–203.

Lee LF. 2007a. Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* **137**: 489 – 514.

Lee LF. 2007b. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* **140**: 333–374.

Leung MP. 2015. Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics* **188**: 182–195.

Manta A, Ho AT, Huynh KP, Jacho-Chavez DT. 2022. Estimating social effects in a multilayered linear-in-means model with network data. *Statistics & Probability Letters* **183**: 109331. ISSN 0167-7152.

Mele A. 2017. A structural model of dense network formation. *Econometrica* **85**: 825–850.

Miguel E, Kremer M. 2004. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* **72**: 159–217.

Miyauchi Y. 2016. Structural estimation of pairwise stable networks with nonnegative externality. *Journal of Econometrics* **195**: 224–235.

Moffitt RA. 2001. Policy interventions, low-level equilibria and social interactions. In Durlauf S, Young P (eds.) *Social Dynamics*. MIT Press, 45–82.

Newey WK, West KD. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**: 703–708.

Nicoletti C, Rabe B. 2019. Sibling spillover effects in school achievement. *Journal of Applied Econometrics* **34**: 482–501.

Nicoletti C, Salvanes KG, Tominey E. 2018. The family peer effect on mothers' labor supply. *American Economic Journal: Applied Economics* **10**: 206–234.

Qu X, Lee LF. 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics* **184**: 209–232.

Sheng S. 2020. A structural econometric analysis of network formation games through subnetworks. *Econometrica* **88**: 1829–1858.

Zacchia P. 2019. Knowledge Spillovers through Networks of Scientists. *The Review of Economic Studies* **87**: 1989–2018.

Social Interactions in Multilayered Observational Networks – Online Appendix –

Juan Estrada

August 15, 2022

Appendix A Bayesian Game of Social Interactions

Players. I consider a set of $\mathcal{I}_N$ individuals who, for simplicity, are characterized by only one individual feature $x_{N,i} \in \mathbb{R}$, instead of Q characteristics, as in section 3. I also assume the existence of an additional characteristic $\zeta_{N,i} \in \mathbb{R}$ that agents observe only at the outset of the first stage. The vector $\boldsymbol{\zeta}_N \in \mathbb{R}^N$ is used to partition individuals into K disjoint clusters $\{C_1, \ldots, C_K\}$, such that the closer individuals in terms of the characteristic $\zeta_{N,i} \in \mathbb{R}$ belong to the same cluster. Finally, individuals are also endowed with $b_{N,i} \in \mathbb{R}$, which is a private characteristic observable only to individual i.

Actions. In the first stage individuals choose a set of connections in the multilayer network $\mathcal{M}_N$. From Assumption 1, the multilayer network has no cycles, $w_{N,m;i,i} = 0$ for all m, and, without loosing generality, it is assumed that the network $\mathcal{M}_N$ is directed for this section. Therefore, in the first stage, individual i chooses actions $\mathbf{w}_{N,m;i} = [w_{N,m;i,1}, \ldots, w_{N,m;i,N}] \in \{0,1\}^N$, for each layer $m \in \{1, \ldots, M\}$. In the second stage, individuals choose an action (outcome) $y_{N,i} \in \mathbb{R}$ that can be interpreted, for example, as an effort level.

Information. The characteristic $\zeta_{N,i} \in \mathbb{R}$ is common knowledge only at the outset of the first stage. Let $\{C_1, \ldots, C_K\}$ be the sequence of K disjoint clusters separating individuals based on their distance in the vector of characteristics $\boldsymbol{\zeta}_N$. Then, (i) individual i only observes x_j if both individuals i and j belong to the same cluster C_k for any arbitrary cluster k. Moreover, individual i observes x_j if they share a connection in any layer, i.e., $i, j \in E_m$ for any $m \in M$, and (ii) for all i, the realization of $b_{N,i}$ is private information. Intuitively, this informational assumption imposes restrictions on individuals' knowledge based on distances along the vector of characteristics $\boldsymbol{\zeta}_N$. The idea of this distance can include, for example, geographical distances, as discussed in Kuersteiner (2019), or economic distances, as in Zacchia (2019).

Distributions there is a prior exogenous distribution of types on T, which is common knowledge, and it is denoted by ρ_b. I also assume the existence of a conditional distribution characterizing the random vector $\mathbf{x}_N$ given $\boldsymbol{\zeta}_N$. This conditional distribution is denoted by $\rho_{x,\zeta}$.The following assumption imposes *convenient* restrictions on $\rho_{x,\zeta}$.

Assumption A.10 (Characterization of ρ_b and $\rho_{x,\zeta}$) *(i) Second moments of ρ_b exist. (ii) The conditional distribution $\rho_{x,\zeta}$ allows for correlation of x_i and x_j if both individuals i and j belong to the same cluster C_k, for some $k \in K$, and they are independent otherwise.*

Equilibrium and Timing. The subgame perfect Nash equilibrium of this two-stage game is characterized by applying backward induction. Therefore, I start by describing the Bayesian social interaction game and characterizing its unique equilibrium. Then, conditional on the equilibrium expected payoffs, I define and characterize the solution of the network formation process.

Appendix A.1 Bayesian Game of Social Interactions

In this game, individuals maximize their expected utility given their beliefs by choosing an action (outcome) $y_{N,i} \in \mathbb{R}$. A quadratic utility function with strategic complementarities is assumed as in Calvó-Armengol et al. (2009). I extend the utility function to include heterogeneous peer and contextual effects for the M possible layers in the multilayer network $\mathcal{M}_N$:

$$\begin{aligned} U_i\left(\mathbf{x}_N, \mathbf{y}_N, \mathcal{M}_N, b_{N,i}\right) = & \left(\gamma^0 x_{N,i} + b_{N,i} + \sum_{m=1}^{M} \sum_{j \neq i} w_{N,m;i,j} x_{N,j} \delta_m^0\right) y_{N,i} \\ & + \sum_{m=1}^{M} \sum_{j \neq i} w_{N,m;i,j} y_{N,i} y_{N,j} \beta_m^0 - \frac{1}{2} y_{N,i}^2. \end{aligned}$$

Note that the informational assumptions guarantee that any individual i can evaluate the value of her utility for all the necessary peers characteristics $x_{N,j}$. As in Blume et al. (2015), I look for a Bayes-Nash equilibrium where individuals maximize their expected payoffs given their types. Individuals' types are constructed as follows. Let $\mathbf{x}_{N,i}$ be the vector of characteristics for all individuals j such that either $i, j \in C_k$ or $i, j \in E_m$ for any $m \in M$, and let N_i be the number of individuals in that vector. Therefore, individual i's type is a vector $[\mathbf{x}_{N,i}, b_{N,i}] \in \mathbb{R}^{N_i+1}$, while the vector of all players' types is $[\mathbf{x}_N, \mathbf{z}_N] \in \mathbb{R}^{2N} \equiv T$. Equilibrium beliefs are constructed from the individuals' strategy functions and the common prior distributions. Formally, an individual's i strategy is a function $s_i : \mathbb{R}^{N_i+1} \to \mathbb{R}$ mapping possible types $[\mathbf{x}_{N,i}, b_{N,i}]$ to actions y_i.

A Bayesian-Nash equilibrium is a vector of strategy profiles $s(\mathbf{x}_N, \mathbf{b}_N)$ where each s_i maximizes $\mathbb{E}[U_i\left(y_{N,i}, y_{N,-i}\right) \mid \mathbf{x}_{N,i}, b_{N,i}]$, and the expectation is taken with respect to s_{-i} and ρ_z (Blume et al., 2015). In addition to Assumption 1, the existence of a Bayesian-Nash equilibrium requires Assumption A.10(i) to hold. This assumption is necessary to ensure that the expected utility is well defined. The second moment is required because of the quadratic form in actions in the strategic complementarities payoffs. The following Proposition characterizes the Bayes-Nash equilibrium of the social interactions game conditional on the multilayer network from the first stage.

Proposition A.2 (Bayes-Nash Equilibrium Social Interactions Game) *If the Bayesian game of social interaction satisfies Assumptions 1 and A.10(i), it has a unique Bayes-Nash equilibrium. The equilibrium strategy profile can be written as*

$$s(\mathbf{x}_N, \mathbf{b}_N) = \left(\mathbf{I}_N - \sum_{m=1}^{M} \beta_m^0 \mathbf{W}_{N,m}\right)^{-1} \left(\gamma^0 x_{N,i} + \sum_{m=1}^{M} \delta_m^0 \mathbf{W}_{N,m} x_{N,j}\right) + \mu(\mathbf{x}_N, \mathbf{b}_N) + \mathbf{b}_N \quad \text{(A-17)}$$

where $\mu_i(\mathbf{x}_N, \mathbf{b}_N)$ depends only on $\mathbf{x}_{N,i}$ and $b_{N,i}$.

Proof. Define $\psi_{N,i} = \gamma x_{N,i} + b_{N,i} + \sum_{m=1}^{M} \sum_{j \neq i} w_{N,m;ij} x_{N,j} \delta_m$, and let $\boldsymbol{\psi}_N = [\psi_{N,1}, \dots \psi_{N,N}]$. With this definition, the utility function in the main text can be written as

$$U_i\left(y_{N,i}, y_{N,-i}\right) = \psi_{N,i} y_{N,i} + \sum_{m=1}^{M} \sum_{j \neq i} w_{N,m;i,j} y_{N,i} y_{N,j} \beta_m - \frac{1}{2} y_{N,i}^2. \quad \text{(A-18)}$$

From Assumption A.10(i), the expected utilities from based on equation (A-18) are well defined. Let $\mathcal{S}$ be the space of all possible vectors of strategy profiles such that $s \in \mathcal{S}$. Endow $\mathcal{S}$ with the L_p^2 max norm, $\|s\| = \max_{i \in \mathcal{I}_N} \|s_i\|_2$, where $\|\cdot\|_2$ is the L^2 norm for a function. Since the strategies are in L_p^2, the expected payoff to any i of any strategy profile is finite (Blume et al., 2015). The first-order conditions for expected utility maximization are that for each i, and given the strategy profile s_{-i} of the other individuals and type $(\mathbf{x}_{N,i}, z_{N,i}) \in \mathcal{T}$,

$$\psi_{N,i} - y_{N,i} + \sum_{m=1}^{M} \sum_{j \neq i} w_{N,m;ij} y_{N,j} \beta_m = 0.$$

Since the utility function is concave in $y_{N,i}$, the first-order conditions are sufficient for expected utility maximization. To prove the existence of an equilibrium, define the mapping $\mathcal{T} : \mathcal{S} \to \mathcal{S}$ such that

$$\mathcal{T}(s) = \boldsymbol{\psi}_N + \sum_{m=1}^{M} \beta_m \mathbf{W}_{N,m} \mathbb{E}[s \mid \mathbf{x}_N, \mathbf{b}_N],$$

where the conditional expectation for individual i with respect to strategy profile of another individual j is given by $\mathbb{E}[s_j \mid \mathbf{x}_{i,N}, z_{N,i}]$ and the conditioning set $(\mathbf{x}_{i,N}, z_{N,i})$ includes all the information available to that individual. A fixed point of $\mathcal{T}$ satisfies the first-order condition for all i, $\psi_{N,i}$, and $z_{N,i}$. Therefore, a fixed point of $\mathcal{T}$ is a Bayes-Nash equilibrium profile. I now show that under Assumption 1, this map is a contraction with contraction constant $\sum_{m=1}^{M} |\beta_m| \|\mathbf{W}_{N,m}\|_\infty$. Let $s' \in \mathcal{S}$ be any arbitrary strategy profile. Then,

$$
\begin{aligned}
\|\mathcal{T}(s)-\mathcal{T}(s')\| &= \left\| \sum_{m=1}^{M} \beta_m \mathbf{W}_{N,m} \mathbb{E}\left[s-s' \mid \mathbf{x}_N, \mathbf{z}_N\right] \right\| \\
&\leq \sum_{m=1}^{M} \left\| \beta_m \mathbf{W}_{N,m} \mathbb{E}[s-s' \mid \mathbf{x}_N, \mathbf{z}_N] \right\| \\
&\leq \sum_{m=1}^{M} |\beta_m| \|\mathbf{W}_{N,m}\|_\infty \mathbb{E}[s-s' \mid \mathbf{x}_N, \mathbf{z}_N] \| \\
&= \left(\sum_{m=1}^{M} |\beta_m| \|\mathbf{W}_{N,m}\|_\infty \right) \left\| \mathbb{E}[s-s' \mid \mathbf{x}_N, \mathbf{z}_N] \right\| \\
&\leq \left\| s-s' \right\|,
\end{aligned}
$$

where the first inequality follows from the triangle inequality; the second inequality follows from dividing the norm between the two maximums $\max_{i \in \mathcal{I}_N} \sum_{j=1}^{N} |w_{N,m;ij}|$ and $\max_{i \in \mathcal{I}_N} \|s_i\|_2$; and the last inequality follows from the contraction constant $\sum_{m=1}^{M} |\beta_m| \|\mathbf{W}_{N,m}\| < 1$, and the fact that the maximum of the expected difference between any two strategy profiles should be not larger than the maximum of the difference of any two strategy profiles. Thus, it follows that $\mathcal{T}$ is a contraction mapping in $(\mathcal{F}, \|\cdot\|)$, and so a fixed point exists and is unique. The rest of the argument follows the proof in Blume et al. (2015). Under the information assumptions, individuals can only use the local information intracluster and across connections to infer any private types. Therefore, $\mu_i(\mathbf{x}_N, \mathbf{z}_N)$ depends only on $\mathbf{x}_{N,i}$ and $z_{N,i}$. ■

Equation (A-17) presents a theoretical foundation for the reduced form equation for the MLiM model in (3). Moreover, the fact that the unobserved component $\mu_i(\mathbf{x}_N, \mathbf{b}_N)$ is correlated with the vector of characteristics $\mathbf{x}_{N,i}$ for each individual i, provides a theoretical argument for the endogeneity issue caused by the network formation. The next step in solving the game is to consider the expected utility of the unique second stage equilibrium as an input for the utility function in the first stage network formation process.

Appendix A.2 Multilayer Network Formation Model

A critical difference between the existing network formation models and the set up I propose is the multilayer network data structure. Following Joshi et al. (2020), I assume a sequential multilayer formation process where individuals choose their set of connections $\mathbf{w}_{N,m;i}$ one layer at a time. Differing from Joshi et al. (2020), for simplicity, I do not assume that the process is repeats exhaustively, but it ends when the players decide connections in the last layer M. This structure generates a sequence of $\{1, \ldots, M\}$ conditional link choices. To further simplify the problem, I assume that individuals behave myopically in the sense that they consider only the structure of the multilayer network at the $m-1$ stage when deciding connections in layer m. To formalize this behavioral assumption into the payoff function, let $\mathbf{W}_{N,m-1}$ be the adjacency matrix of the multilayer that has emerged up to layer $m-1$, where $w_{N,m-1;i,j} = 1$ if individuals i and j have a link in any of the layers in the sequence of layers $\{1, 2, \ldots, m-1\}$. Finally, payoff externalities (Miyauchi, 2016; Mele, 2017; De Paula et al., 2018; Christakis et al., 2020; Sheng, 2020), or degree heterogeneity (Graham, 2017) are ruled out for the sake of this characterization. However, this assumption is imposed only for simplicity, and the model could be extended to have payoff externalities and degree heterogeneity as long as there exist an equilibrium selection that chooses the multilayer network realized in the second stage when there are multiple equilibria, see Leung (2015) for the incomplete information setting.

To complete the characterization of the payoff function, I assume that there is an additional utility component such that individuals experience higher marginal utilities when they connect otherwise disconnected components in different clusters. This relevant feature where social structures are clusters of dense connection linked by occasional relations between groups is known in the literature as bridging structural holes (Burt, 2004). Let individual i be part of an arbitrary cluster k. Therefore, individual i's utility for layer m is given by

$$\begin{aligned} U_{i,m}\left(\mathbf{x}_i, \mathbf{v}_i, \mathbf{W}_{N,m-1}\right) = \sum_{j=1}^{N} w_{N,m;i,j} (\pi_{x,1} |x_{N,i} - \mathbb{E}[x_{N,j} \mid \boldsymbol{\zeta}_N, \mathbf{x}_{N,i}, b_{N,i}]| + \pi_{x,2} \mathbb{E}[x_{N,j} \mid \boldsymbol{\zeta}_N, \mathbf{x}_{N,i}, b_{N,i}] \\ + \pi_b \mathbb{E}[b_{N,j} \mid \boldsymbol{\zeta}_N, \mathbf{x}_{N,i}, b_{N,i}] + \pi_C \mathbf{1}\{j \in C_\ell\} \mathbf{1}\{i \in C_k\} \sum_i w_{N,m-1;i,j} - v_{N,m;i,j}), \end{aligned} \tag{A-19}$$

where $v_{N,m;i,j}$ are mutually independent and identically distributed with marginal distribution ρ_v, for all $i, j \in \mathcal{I}_N$ and $m \in \{1, \ldots, M\}$. From Assumption A.10(i), it follows that $\mathbb{E}[x_{N,j} \mid \boldsymbol{\zeta}_N, \mathbf{x}_{N,i}, b_{N,i}] = x_{N,j}$ if individuals i and j belong to the same cluster or if they are connected in any layer of the multilayer network. The additional components $|x_{N,i} - \mathbb{E}[x_{N,j} \mid \boldsymbol{\zeta}_N, \mathbf{x}_{N,i}, b_{N,i}]|$, $\mathbb{E}[x_{N,j} \mid \boldsymbol{\zeta}_N, \mathbf{x}_{N,i}, b_{N,i}]$, and $\mathbb{E}[b_{N,j} \mid \boldsymbol{\zeta}_N, \mathbf{x}_{N,i}, b_{N,i}]$ represent a homophily component based on the characteristics of individuals i and j, and the effect on the expected utility that a connection with j will bring individual i in the second stage via the characteristics $x_{N,j}$ and $b_{N,j}$. The clusters k and ℓ are such that $k \neq \ell$ and $\mathbf{1}\{\cdot\}$ represents the indicator function. The component $\mathbf{1}\{j \in C_\ell\} \mathbf{1}\{i \in C_k\} \sum_j w_{N,m-1;i,j}$ captures the idea of bridging structural holes. Individual i's utility of creating a connection with j on layer m increases if j has more connections in layers $\{1, 2, \ldots, m-1\}$, and they belong to different clusters $\in C_k$ and $\in C_\ell$.

Given the simplifying assumption that individuals' payoffs do not depend on the structure of the contemporaneous network and they make choices myopically, the optimization problem of choosing connections to maximize utility for individual i does not depend on other individuals' decisions. Therefore, conditional on the realizations of the shocks $v_{N,m;i,j}$, the characteristics $x_{N,i}$ and $b_{N,i}$, and the conditional expectations taken with respect to the common knowledge distributions ρ_b and $\rho_{x,\zeta}$, the solution for the multilayer network formation process exists and is unique. Models without network externalities which produce unique solutions have been explored in the dyadic regression literature (Graham, 2017, 2020).

The next step is to impose restrictions on the distributions ρ_b, $\rho_{x,\zeta}$ and ρ_b that are sufficient to producing multilayer network solutions where the dependence between individuals' characteristics disappears after a finite path length or a finite number of edge type changes. Define p_{in}^{min} and p_{out}^{max} as the *minimum* and *maximum* probabilities induced by the distributions ρ_b, $\rho_{x,\zeta}$, and ρ_v that an individual forms a connection with another *inside* and *outside* her cluster, respectively. Additionally, let N_k be the number of individuals in cluster $k \in \{1, \ldots, K\}$. The following assumptions are imposed on these objects.

Assumption A.11 (Linking Probabilities) *The conditional distribution $\rho_{x,\zeta}$ and the probability distribution ρ_b and ρ_v are such that for some constant c and any $k \in \{1, \ldots, K\}$, (i) $p_{in}^{min} \geq c \log N_k / N_K$, (ii) $p_{out}^{max} \geq c \log K/K$, and (iii)*

$$\left\lceil \frac{\log\left(33c^2/400\right) N_k \log N_k)}{\log(N_k p_{in}^{min})} \right\rceil + 2\left\lfloor \frac{1}{c} \right\rfloor + 2 < \left\lceil \frac{\log(cK/11)}{\log(K p_{out}^{max})} \right\rceil.$$

Assumptions A.11 *(i)* and *(ii)* impose restrictions on the minimum and maximum probabilities of links within and across clusters. Part *(i)* guarantees that the diameter of the subnetworks inside an arbitrary cluster k will be bounded by the left-hand side of the inequality on part *(iii)* for the case of a constant probability of connection given by p_{in}^{min}. For the connections across clusters, I consider an auxiliary network where each cluster $k \in \{1, \ldots, K\}$ represents a node, and connections are formed with a constant probability p_{out}^{max}. Therefore, part *(ii)* guarantees that the diameter of the auxiliary inter-cluster network has a lower bound given by the right-hand-side of the inequality on part *(iii)*. These results follow from Theorem 4 in Chung and Lu (2001), and can be extended to weighted random graphs using the results in Amini and Lelarge (2015). The next lemma follows directly from Assumption A.11. Importantly, these results represent worst-case-scenarios using a *minimax* approach where the largest possible intra-cluster upper bound is compared with the smallest possible inter-cluster lower bound.

Lemma 1 (Diameter Comparison) *Let individuals $i, j \in C_k \subset \mathcal{I}_N$ have the largest shortest path length for an arbitrary $k \in \{1, \ldots, K\}$. If Assumption A.11 holds, then, there exists two individuals $r, s \in \mathcal{I}_N$ such that $r \in C_k$ and $s \notin C_k$ for which their shortest path length is larger than that between individuals i and j.*

From Lemma 1 it also follows that there exist a path connecting two individuals from different clusters that will have at most as many edge type changes than any path connecting two individuals in the same cluster. It is crucial to highlight that condition *(iii)* in Assumption A.11 is a worst-case scenario because it assumes that only one edge connects any two clusters. Given that each cluster can have more than one individual, there is likely more than one connection inside each cluster in the path connecting them, making the assumption easier to fulfill.

The conclusion that there exists at least one inter-cluster path with larger geodesic distance and edge type changes than any intra-cluster path is fundamental to characterizing the constants from proposition 1. Under this network formation model, I can define the constants K_c and K_d as the minimum inter-cluster number of edge type changes ($c^*_N(i,j)$) and geodesic distance ($d^*_N(i,j)$), respectively, larger than the maximum values for those quantities considering only intra-cluster connections. The following proposition formalizes the previous discussion.

Proposition A.3 (Statistical Properties Social Interactions Game) *Let Assumptions A.10 and A.11 hold. Define K_c and K_d as the minimum number of edge type changes and geodesic distance, respectively, larger than the maximum values for those quantities considering only inter-cluster links. Then, if for any i and j, $c^*_N(i,j)$ and $d^*_N(i,j)$ are such that $c^*_N(i,j) \geq K_c$ or $d^*_N(i,j) \geq K_d$, then $\mu_i(\mathbf{x}_N, \mathbf{b}_N)$ in equation (A-17) is independent of x_j.*

This proposition follows directly from the statistical and informational assumptions discussed in this section. The main implication of proposition A.3 is that, conditional on the equilibrium multilayer network, the observed characteristics of an individual j who is far away in the multilayer network space are not correlated with the unobserved characteristics of individual i. This result motivates the WND assumption, which justifies the use of the conditional moments (6) and (7) as a source for identification. In the context of Blume et al. (2015), the main insight driving the result, is that an individuals i cannot use the characteristics x_j to predict the private types when they do not belong to the same cluster. The following section formalizes the identification result.

Appendix B Auxiliary Results

Lemma 2 (Invertibility of $\mathbf{S}(\boldsymbol{\beta}, \mathcal{M}_N)$) *Let Assumption 1 hold, then $\mathbf{S}(\boldsymbol{\beta}, \mathcal{M}_N) = \mathbf{I}_N - \sum_{m=1}^M \beta_m \mathbf{W}_{N,m}$ is invertible.*

Proof. In the trivial case where $\beta_m = 0$ for all m, it follows that $\mathbf{S}(\beta, \mathcal{M}_N) = \mathbf{I}_N$ which is inevitable. For the non-trivial case, let $\theta = \sum_{m=1}^M |\beta_m|$ and note that $\theta = 0$ if only if $\beta_m = 0$ for all m. For $\theta \neq 0$, $\mathbf{S}(\boldsymbol{\beta}, \mathcal{M}_N)$ can be written as

$$\mathbf{S}(\boldsymbol{\beta}, \mathcal{M}_N) = \mathbf{I} - \theta \left(\frac{1}{\theta} \sum_{m=1}^{m} \beta_m \mathbf{W}_{N,m} \right) \equiv I - \theta \mathbf{A},$$

where $\mathbf{A} \equiv 1/\theta \sum_{m=1}^m \beta_m \mathbf{W}_{N,m}$. To show that $\mathbf{S}(\boldsymbol{\beta})$ has an inverse, it is enough to show that $\det(I - \theta \mathbf{A}) \neq 0$. Note that the Gerschgorin's disk of $\mathbf{A}$ is given by $R_i = \sum_{j=1, i \neq j}^N |a_{i,j}|$. By assumption 1, all the matrices forming $\mathbf{A}$ have zeros in the main diagonal, thus $R_i = \sum_{j=1, i \neq j}^N |a_{i,j}| = \sum_{j=1}^N |a_{i,j}|$. Let λ_i be the ith eigenvalue of $\mathbf{A}$, then by Gerschgorin's (1931) Circe Theorem, λ_i lies within at least one of the Gershgorin discs centered in zero with radius R_i. Given that all the circles are centered in zero, it follows that $|\lambda_i| \leq \sup_i \sum_{j=1}^N |a_{i,j}| = \|\mathbf{A}\|_\infty$ for all i. Note that $\sup_i \sum_{j=1}^N |a_{i,j}| =$

$\sup_i \sum_{j=1}^N \left|\frac{1}{\theta}\sum_{m=1}^M \beta_m w_{N,m;i,j}\right| \leq \left|\frac{1}{\theta}\right| \sum_{m=1}^M |\beta_m| \sup_i \sum_{j=1}^N |w_{N,m;i,j}| = \left|\frac{1}{\theta}\right| \sum_{m=1}^M |\beta_m| \, \|\mathbf{W}_{N,m}\|_\infty$, where the second inequality follows from the triangle inequality and the supremum of the sum being at most the sum of the supremum. Therefore, $|\lambda_i| \leq |1/\theta| \sum_{m=1}^M \beta_m \|\mathbf{W}_{N,m}\|_\infty$. Note that if λ_i is an eigenvalue of $\mathbf{A}$, then $(1-\theta\lambda_i)$ is an eigenvalue of $I-\theta\mathbf{A}$. Moreover, given that $I-\theta\mathbf{A}$ is a $N\times N$ matrix, its determinant is given by the product of its eigenvalues, i.e., $\det(I-\theta\mathbf{A}) = \Pi_i(1-\theta\lambda_i)$. From the discussion before, $|\theta\lambda_i| \leq \sum_{m=1}^M \beta_m \|\mathbf{W}_{N,m}\|_\infty < 1$ for all i, where the second inequality comes from Assumption 1. Thus, $\Pi_i(1-\theta\lambda_i) \neq 0$, completing the proof. ■

Remark 2 (Series Expansion) Assumption 1 also guarantees that it is possible to write the matrix $\mathbf{S}(\boldsymbol{\beta}, \mathcal{M}_N)$ as an infinite series given by $\mathbf{S}(\boldsymbol{\beta}, \mathcal{M}_N) = \sum_{r=0}^\infty \left(\sum_{m=1}^M \beta_m \mathbf{W}_{N,m}\right)^r$. Let the matrix $\mathbf{A}_N(\zeta,\phi)$ represent a product of ζ adjacency matrices containing a possible combination of edge types given by the sequence ϕ. This notation simplifies the representation of large multiplications of adjacency matrices. For instance, for any arbitrary layers $m_1, m_2, m_3 \in \{1,\ldots,M\}$, the matrix $\mathbf{W}_{N,m_1}\mathbf{W}_{N,m_2}\mathbf{W}_{N,m_3}$ can be represented by $\mathbf{A}_N(3, \{m_1,m_2,m_3\})$. With this additional notation, it is possible to explicitly write the solution for equation (2) in the main text as an infinite sum of the product of different adjacency matrices given by

$$
\begin{aligned}
\mathbf{y}_N = & \sum_{\zeta=0}^{\infty} \sum_{\phi\in\mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell\in\phi}\beta_\ell\right) \sum_{m=1}^M \mathbf{A}_N(\zeta,\phi)\mathbf{W}_{N,m}\mathbf{X}_N\boldsymbol{\delta} \\
& + \sum_{\zeta=0}^{\infty} \sum_{\phi\in\mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell\in\phi}\beta_\ell\right) \mathbf{A}_N(\zeta,\phi)\widetilde{\mathbf{X}}_N\widetilde{\boldsymbol{\gamma}} + \sum_{\zeta=0}^{\infty} \sum_{\phi\in\mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell\in\phi}\beta_\ell\right) \mathbf{A}_N(\zeta,\phi)\boldsymbol{\varepsilon}_N,
\end{aligned}
\tag{B-20}
$$

where $\mathbb{P}\big(\{1,\ldots,M\},\zeta\big)$ represents the sequence of all possible ζ-permutations with repetition from the set of possible layers $\{1,\ldots,M\}$. By convention, the matrix $\mathbf{A}_N(0,0) = \mathbf{I}_N$ and $\beta_0 = 1$. Importantly, note that the (i,j)th element of the matrix $\mathbf{W}_{N,m}^k$ gives the number of paths of length k from agents i to j (for some layer m), see e.g., Graham (2015), while the (i,j)th element of the product of two adjacency matrices $\mathbf{W}_{N,m}$ and $\mathbf{W}_{N,s}$ for layers m and s, contains the number of paths of length two between nodes i and j where each path begins with a type m edge and changes to type s after the second node in the sequence. In general, the (i,j)th position of the matrix formed by k products of adjacency matrices from different layers gives the number of k-paths between individuals i and j that change edge types k times. Therefore, (B-20) shows that both interlayer and intralayer indirect connections can be used as relevant instruments for $\mathbf{W}_{N,m}\mathbf{y}_N$. The required necessary conditions on the parameters to guarantee instruments relevance are not straightforward to characterize analytically. The reason is that many coefficients associated with the potential instruments involve complicated non-linear functions of the structural parameters. However, it is possible to provide sufficient conditions to guarantee the relevance of a subset of possible instruments. Following Manta et al. (2022), it is possible to rewrite (B-20) as

$$
\begin{aligned}
\mathbf{y}_N =& \alpha \sum_{r=0}^{\infty} \left(\sum_{m=1}^{M} \beta_m \mathbf{W}_{N,m} \right)^r \iota + \sum_{q=1}^{Q} \gamma_q \mathbf{x}_N^q + \sum_{q=1}^{Q} \sum_{m=1}^{M} \left(\sum_{r=0}^{\infty} \beta_m^r \mathbf{W}_{N,m}^{r+1} \right) \mathbf{x}_N^q \left(\delta_{q;m} + \gamma_q \beta_m \right) \\
&+ \sum_{q=1}^{Q} \sum_{m \neq s \in \{1,\ldots M\}} \left(\sum_{r=0}^{\infty} \sum_{r'=0}^{\infty} \beta_m^r \beta_s^{r'} \mathbf{W}_m^{r+1} \mathbf{W}_s^{r'+1} \right) \mathbf{x}_N^q \left(\delta_{q;m} + \gamma_q \beta_m \right) + \ldots + \\
&+ \sum_{r=0}^{\infty} \left(\sum_{m=1}^{M} \beta_m \mathbf{W}_m \right)^r \boldsymbol{\varepsilon}_N,
\end{aligned} \tag{B-21}
$$

where the dots notation masks the additional products between the different adjacency matrices. Note that the instruments formed by the product of infinite powers of any two adjacency matrices m and s are relevant so long as $\max_{q \in \{1,\ldots,Q\}} \left(|\gamma_q \beta_1 + \delta_{q;1}|, \ldots, |\gamma_q \beta_M + \delta_{q;M}| \right) \neq 0$.

Remark 3 (Solution for the Outcome Equation) Equation (B-21) in Remark 2 shows that the outcome equation for $\mathbf{y}_N$ can be written as an infinite sum of the adjacency matrices' products. All multiplications between unweighted adjacency matrices produce new matrices with the property that their (i,j)th entry contains the number of paths at a certain distance and with a certain number of edge changes between individuals i and j. For instance, the (i,j)th position of the matrix $\mathbf{W}_m^k$ contains the number of length-k paths from agents i to j in layer m. The (i,j)th position in the matrix $\mathbf{W}_m \mathbf{W}_s$ gives the number of length 2 paths that start with an edge-type m and changes to an edge-type s. The i,jth position in the matrix $\mathbf{W}_m \mathbf{W}_s \mathbf{W}_r$ gives the number of length 3 paths that start with an edge-type m, changes to an edge-type s, and changes again to an edge-type r. The pattern can be extended to any possible combination of products between the matrices $\mathbf{W}_m$ for $m \in \{1,\ldots,M\}$.

The empirical counterpart of the outcome equation for $\mathbf{y}_n$ can be decomposed into three possible types of summands: (1) $\mathbf{A}_n(\zeta,\phi)\mathbf{W}_{n,m}\mathbf{X}_n$, (2) $\mathbf{A}_n(\zeta,\phi)\widetilde{\mathbf{X}}_n$, and (3) $\mathbf{A}_n(\zeta,\phi)\boldsymbol{\varepsilon}$ for $m \in \{q,\ldots,M\}$, where $\mathbf{A}_n(\zeta,\phi)$ was introduced in Remark 2 for Lemma 2. As mentioned before, this notation facilitates to write the products of adjacency matrices. From the example above, the matrix $\mathbf{W}_m\mathbf{W}_s\mathbf{W}_r$ can be represented by $\mathbf{A}_n(3,\{m,s,r\})$. Similarly, the matrix $\mathbf{W}_m^k$ can be written as $\mathbf{A}(k,\{m,\ldots,m\})_n$, where the sequence $\{m,\ldots,m\}$ contains k elements. Denoting $\mathbb{P}\big(\{1,\ldots,M\},\zeta\big)$ as in Remark 2, it follows that the sample analogue of equation (B-20) can be written as

$$
\begin{aligned}
\mathbf{y}_n =& \sum_{\zeta=0}^{\infty} \sum_{\phi \in \mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell \in \phi} \beta_\ell \right) \sum_{m=1}^{M} \mathbf{A}_n(\zeta,\phi) \mathbf{W}_{n,m} \mathbf{X}_n \boldsymbol{\delta} \\
&+ \sum_{\zeta=0}^{\infty} \sum_{\phi \in \mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell \in \phi} \beta_\ell \right) \mathbf{A}_n(\zeta,\phi) \widetilde{\mathbf{X}}_n \widetilde{\boldsymbol{\gamma}} + \sum_{\zeta=0}^{\infty} \sum_{\phi \in \mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell \in \phi} \beta_\ell \right) \mathbf{A}_n(\zeta,\phi) \boldsymbol{\varepsilon}_n.
\end{aligned} \tag{B-22}
$$

Lemma 3 *Let Assumptions 2 hold for* $\{\mathbf{r}_{n,i}\}_{n\geq 1}$, $i \in \mathcal{I}_n$. *Define* $R_{i,j} = f_{q,\ell}(\mathbf{r}_{n,\{i,j\}}) = r_{n,i,q}r_{n,j,\ell}$ *and* $R_{h,s} = g_{q',\ell'}(\mathbf{r}_{n,\{h,s\}}) = r_{n,h,q'}r_{n,s,\ell'}$ *for* $i, j, h, s \in \mathcal{I}_n$, *where* q, q', ℓ, *and* ℓ' *are components of the vector* $\mathbf{r}_{n,i}$. *Let Assumption 6 hold for* $R_{i,j}$ *and* $R_{h,s}$. *Then*

$$|\mathrm{Cov}\,(R_{i,j}, R_{h,s})| \leq 2\bar{\theta}_{n,s}(C+16) \times 4\,(\pi_1 + \tilde{\gamma}_1)\,(\pi_2 + \tilde{\gamma}_2)\,\underline{\theta}_{n,s}^{1-p_f-p_g}, \tag{B-23}$$

where $\underline{\theta}_{n,s} = \theta_{n,s} \wedge 1$, $\bar{\theta}_{n,s} = \theta_{n,s} \vee 1$, $\pi_1 = \|\mathbf{r}_{n,i}\|_{p_{f,i}}\|\mathbf{r}_{n,j}\|_{p_{f,j}}$, $\pi_2 = \|\mathbf{r}_{n,h}\|_{p_{f,h}}\|\mathbf{r}_{n,s}\|_{p_{f,s}}$, $\widetilde{\gamma}_1 = \max\{\|\mathbf{r}_{n,i}\|_{p_{f,i}+p_{f,j}}, \|\mathbf{r}_{n,j}\|_{p_{f,i}+p_{f,j}}\}$, $\widetilde{\gamma}_2 = \max\{\|\mathbf{r}_{n,h}\|_{p_f}, \|\mathbf{r}_{n,s}\|_{p_g}\}$ *where* $p_f = 1/p_{f,i} + 1/p_{f,j}$ *and* $p_g = 1/p_{g,h} + 1/p_{g,s}$. *The constant* C *is the same as in Assumption 2, the indexes* i, j, h, s, *and components* q, q', ℓ *and* ℓ' *may or may not be the same.*

Proof. Define the increasing continuous functions $h_1(x)$ and $h_2(x)$ in (Appendix A in Kojevnikov et al., 2020, pp. 899-907) Theorem A.2 to be $h_1(x) = h_2(x) = x$. Note that the functions $f_{q,\ell}$ and $g_{q',\ell'}$ are continuous, and their truncated version of the form $\varphi_{K_1} \circ f \circ \varphi_{h_1}(K_2)$ and $\varphi_{K_1} \circ g \circ \varphi_{h_1}(K_2)$ for all $K \in (0,\infty)^2$ are in $\mathscr{L}_{Q+1,2}$. Assumption 6 guarantees the existence of the moments defining $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$. Then, Theorem A.2 in (Appendix A in Kojevnikov et al., 2020, pp. 899-907) applies to this setting (see also Corollary A.2.in Appendix Kojevnikov et al., 2020, pp. 899-907). ∎

Lemma 4 (LLN for Products of ψ-dependent Random Variables) *Let Assumptions 2, 6, 7 and 8 hold. Define* $R_{n,i,j} = r_{n,i,q}r_{n,j,\ell}$ *and form* $\{R_{n,i,j}\}_{i\in\mathcal{I}_n, j\in\mathcal{I}_i}$, *where* $\mathcal{I}_n$ *is the set of all individuals in the sample of size* n, *while* $\mathcal{I}_i$ *is a set of indexes define for each* $i \in \mathcal{I}_n$. *Defining the weights* $w_{i,j} \in [0,1]$, *as* $n \to \infty$

$$\left\| \frac{1}{n}\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}\left(R_{n,i,j} - \mathbb{E}\left[R_{n,i,j}\right]\right)\right\|_1 \xrightarrow{a.s.} 0$$

Proof. For simplicity, in this proof, it is assumed that $\mathrm{Var}(R_{n,i,j}) \leq a$ for all i and j, and a generic constant a. However, this assumption is not necessary. Without the finite variance assumption, the proof proceeds similarly but separating $R_{n,i,j} = R^k_{n,i,j} + \widetilde{R}^k_{n,i,j}$, where $\widetilde{R}^k_{n,i,j} = \varphi_k(R_{n,i,j})$, and $\varphi_k(x) = (-K) \vee (K \wedge x_i)$ is a censoring function, see Jenish and Prucha's (2009) proof of Theorem 3 for more details, and Kojevnikov et al.'s (2020) proof for Theorem 3.1. Define the k-norm for a random variable X as $\|X\|_k = (\mathbb{E}[|X|^k)^{1/k}$ for $k \in [1,\infty)$. Thus, by Lyapunov's inequality and the definition of the k-norm it follows that

$$\left\| \frac{1}{n}\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}\left(R_{n,i,j} - \mathbb{E}\left[R_{n,i,j}\right]\right)\right\|_1 \leq \left\| \frac{1}{n}\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}\left(R_{n,i,j} - \mathbb{E}\left[R_{n,i,j}\right]\right)\right\|_2 \tag{B-24}$$

where (B-24) is an expression for the standard deviation of $\sum_{i\in\mathcal{I}_{n,m,\lambda}}\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}$ normalized by the sample size n. Moreover, note that the variance of that quantity can be written as

$$\text{Var}\left(\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}\right) = \sum_{i\in\mathcal{I}_n}\text{Var}\left(\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}\right) + \sum_{i\neq h\in\mathcal{I}_n}\text{Cov}\left(\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}, \sum_{s\in\mathcal{I}_h} w_{h,s}R_{n,h,s}\right)$$

Let us start by analyzing first the variance component. Given the network dependence between individuals, it follows that

$$\begin{aligned}
\text{Var}\left(\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}\right) &= \sum_{j\in\mathcal{I}_i} w_{i,j}^2\text{Var}(R_{n,i,j}) + \sum_{j\neq s\in\mathcal{I}_i} w_{i,j}w_{i,s}\text{Cov}(R_{n,i,j}, R_{n,i,s}) \\
&\leq a\sum_{j\in\mathcal{I}_i} w_{i,j}^2 + \sum_{j\in\mathcal{I}_i}\sum_{d\geq 1}\sum_{s\in\mathcal{P}_n(j,d)\cap\mathcal{I}_i} |\text{Cov}(R_{n,i,j}R_{n,i,s})| \\
&\leq a\sum_{j\in\mathcal{I}_i} w_{i,j}^2 + b\sum_{d\geq 1}\theta_{n,d}\sum_{j\in\mathcal{I}_i} |\mathcal{P}_n(j,d)|,
\end{aligned} \tag{B-25}$$

where the second inequality follows from $w_{i,j}, w_{i,s} \in [0,1]$, and a is a generic constant that follows from the assumption that $\text{Var}(R_{n,i,j})$ is finite (or the fact that after partitioning $R_{n,i,j}$ it is possible to bound its variance). The third inequality comes from the fact that under Assumptions 2 and 6, and Lemma 3, the covariances are bounded by $|\text{Cov}(R_{n,i,j}, R_{i,s})| \leq b\theta_{n,d}$, where $b < \infty$ contains the constants from Lemma 3, and includes the possibility that $\theta_{n,d}$ could have exponents of either 1 or $1-p_f-p_g$. Focusing now on the covariance component of equation (B-24), by the properties of the covariance, the expression inside the summation can be written as

$$\begin{aligned}
\text{Cov}\left(\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}, \sum_{s\in\mathcal{I}_h} w_{h,s}R_{n,h,s}\right) &= \sum_{j\in\mathcal{I}_i}\sum_{s\in\mathcal{I}_h} w_{i,j}w_{h,s}\text{Cov}(R_{n,i,j}, R_{n,h,s}) \\
&\leq \sum_{j\in\mathcal{I}_i}\sum_{d\geq 1}\sum_{s\in\mathcal{P}_n(j,d)\cap\mathcal{I}_h} |\text{cov}(R_{n,i,j}R_{n,h,s})| \\
&\leq b\sum_{d\geq 1}\theta_{n,d}\sum_{j\in\mathcal{I}_i} |\mathcal{P}_n(j,d)|,
\end{aligned} \tag{B-26}$$

where the inequalities in (B-26) follow from the same arguments discussed before. It follows from equations and (B-26) that the total variance can be bounded by

$$\begin{aligned}
\text{Var}\left(\sum_{i\in\mathcal{I}_{n,m,\lambda}}\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}\right) &\leq a\sum_{j\in\mathcal{I}_i} w_{i,j}^2 + 2b\sum_{d\geq 1}\theta_{n,d}\sum_{j\in\mathcal{I}_i} |\mathcal{P}_n(j,d)| \\
&= a\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}^2 + 2b\sum_{d\geq 1}\theta_{n,d}\sum_{i\in\mathcal{I}_n} |\mathcal{P}_n(j,d)| \\
&\leq n\left(a\sum_{i\in\mathcal{I}_n} n^{-1}\sum_{j\in\mathcal{I}_i} w_{i,j} + 2b\sum_{d\geq 1}\theta_{n,d}\bar{D}_n(d)\right),
\end{aligned} \tag{B-27}$$

where the last inequality follows from $w_{i,j} \in [0,1]$. Define $\bar{\mathcal{I}}_i^w = n^{-1}\sum_{j\in\mathcal{I}_i} w_{i,j}$. It follows that combining equations B-24 and B-27,

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}\left(R_{n,i,j}-\mathbb{E}\left[R_{n,i,j}\right]\right)\right\|_1 \le \left(\frac{a}{n}\sum_{i\in\mathcal{I}_n}\bar{\mathcal{I}}_i^w + \frac{2b}{n}\sum_{d\ge 1}\theta_{n,d}\bar{D}_n(d)\right)^{1/2} \tag{B-28}$$

Depending on the component of interest from the matrix $\mathbf{Z}_n^\top\mathbf{D}_n$ and the vector $\mathbf{Z}_n^\top\boldsymbol{\varepsilon}$, the set $\mathcal{I}_i$ can be either: (1) $\mathcal{I}_i = \varnothing$, (2) $\mathcal{I}_i = \mathcal{P}(i,1,m,\lambda)\times\mathcal{P}(i,1,s)$ where $\mathcal{P}(i,1,m,\lambda)$ is the set of individual i's neighbors in the implicit network formed by the weighted adjacency matrix $\mathcal{W}_{n,m,\lambda}$ (set of nodes that can be used to for moment conditions for i), and $\mathcal{P}(i,1,s)$ is the set of i's neighbors in layer s, (3) $\mathcal{I}_i = \mathcal{P}(i,1,m,\lambda)$, or (4) $\mathcal{P}(i,1,m)$. For any of the four cases, Assumption 7 guarantees that $\bar{\mathcal{I}}_i^w = o_{a.s.}(1)$ for all i, and by the algebra of stochastic orders $\sum_{i\in\mathcal{I}_n}\bar{\mathcal{I}}_i^w = o_{a.s.}(1)$. Moreover, Assumption 8 guarantees that $n^{-1}\sum_{d\ge 1}\theta_{n,d}\bar{D}_n(d) \xrightarrow{\text{a.s.}} 0$. It follows that the right hand side of equation (B-28) is $o_{a.s.}(1)$, completing the proof. ■

Lemma 5 (LLN for Outcomes and Regressors) *Let Assumptions 1, 2, 6, 7 and 8 hold hold. Define $\mathbf{x}_{n,q}$ as the qth column of the matrix $\mathbf{X}_n$. As in the main text $\mathbf{w}_{n,m,\lambda,i}$ and $\mathbf{w}_{n,s,i}$ represent the ith row of the matrices $\mathcal{W}_{n,m,\lambda}$ and $\mathbf{W}_{n,s}$, respectively. Then,*

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\mathbf{w}_{n,m,\lambda,i}\mathbf{x}_{n,q}\mathbf{w}_{n,s,i}\mathbf{y}_n - \mathbb{E}[\mathbf{w}_{N,m,\lambda,i}\mathbf{x}_{N,q}\mathbf{w}_{N,s,i}\mathbf{y}]\right\|_1 \xrightarrow{a.s.} 0 \tag{B-29}$$

and

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\mathbf{w}_{n,m,i}\mathbf{y}_n\mathbf{x}_{n,q} - \mathbb{E}[\mathbf{w}_{N,m,i}\mathbf{y}x_{N,q,i}]\right\|_1 \xrightarrow{a.s.} 0 \tag{B-30}$$

Proof. From Assumption 1, as shown in Remark 3, it follows that for any three arbitrary layers m_1, m_2, and m_3 from $\{1,\dots,M\}$, two arbitrary characteristics q and l from $\{1\dots,Q\}$, an arbitrary number of products of adjacency matrices ζ, and an arbitrary sequence ϕ, one single summation from the first component $\mathbf{A}(\zeta,\phi)_n\mathbf{W}_{n,m_2}\mathbf{X}_n$ of the vector $(\mathcal{W}_{n,m_1,\lambda}\mathbf{x}_{n,q})^\top\mathbf{W}_{n,m_2}\mathbf{y}_n$ can be written as:

$$\begin{aligned}\left(\prod_{\ell\in\phi}\beta_\ell\right)&(\mathcal{W}_{n,m_1,\lambda}\mathbf{x}_{n,q})^\top\mathbf{W}_{n,m_2}\mathbf{A}(\zeta,\phi)\mathbf{W}_{n,m_3}\mathbf{x}_{n,l}\delta_{m_3,l}\\ &= \left(\prod_{\ell\in\phi}\beta_\ell\right)(\mathcal{W}_{n,m_1,\lambda}\mathbf{x}_{n,q})^\top\mathbf{A}(\zeta+2,\{m_2,\phi,m_3\})\mathbf{x}_{n,l}\delta_{m_3,l},\end{aligned} \tag{B-31}$$

where the equality follows by the definition of $\mathbf{A}(\zeta,\phi)_n$. For simplicity, define $\mathbf{A}(\zeta+2,\{m_2,\phi,m_3\})_n \equiv \mathbf{A}(\zeta',\phi')_n$. Let $\mathcal{I}_i = \eta_{i,\lambda,m}\times\eta_{i,\zeta',\phi'}$ represent the Cartesian product of $\eta_{i,\lambda,m}$ and $\eta_{i,\zeta',\phi'}$, which are the set of individual i's neighbors in the implicit networks induced by $\mathcal{W}_{n,m,\lambda}$ and $\mathbf{A}(\zeta', \phi')_n$, respectively. Therefore, the right hand side of equation (B-31) can be written as

$$\frac{1}{n} \sum_{\ell \in \mathcal{I}_n} \sum_{i,j \in \mathcal{I}_\ell} w_{i,j} R_{n,i,j}, \tag{B-32}$$

where $w_{i,j} = w_{\lambda;\ell,i} w_{\ell,j}$ and $R_{n,i,j} = x_{n,q} x_{n,l}$. Thus, from Lemma 4, it follows that (B-32) converges to the population expectation. Because the characteristics q and l, the number of products of adjacency matrices ζ, and the sequence ϕ were chosen arbitrarily, this convergence process applies for all the components in the infinite sum in ζ. Given that each component of the sum converges to a finite expectation, the infinite sum of finite expectations is also finite given the restriction on the parameters β_m from Assumption 1. This completes the proof for the convergence of the first component of the outcome equation $\mathbf{y}_n$ in (B-22).

For the second component in (B-22), given by $\mathbf{A}(\zeta, \phi)_n \widetilde{\mathbf{X}}_n$, the proof works analogously by substituting $\mathbf{A}(\zeta + 2, \{m_2, \phi, m_3\})_n$ for $\mathbf{A}(\zeta + 1, \{m_2, \phi\})_n$. Finally, for the third component in (B-22) given by $\mathbf{A}(\zeta, \phi)_n \boldsymbol{\varepsilon}_n$, because $\varepsilon_{n,i}$ is just another component of $\mathbf{r}_{n,i}$ for all i, the results in Lemma 4 also applies for this case. This completes the proof for equation (B-29). The proof for (B-30) is analogous. ∎

Lemma 6 (LLN for Instruments and Regressors) *Let Assumptions 2, 6, 7 and 8 hold. Then as* $n \to \infty$

$$\left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \mathbf{z}_{n,i} \mathbf{d}_{n,i}^\top - \mathbb{E}[\mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top] \right\|_1 \xrightarrow{a.s.} 0. \tag{B-33}$$

Proof. There are four different types of components in the matrix $\mathbf{Z}_n^\top \mathbf{D}_n$ formed by summations of products of (1) non-network regressors of the form $x_{n,i,q} x_{n,i,\ell}$, (2) network regressors of the form $\mathbf{w}_{n,m,\lambda,i} \mathbf{x}_{n,q} \mathbf{w}_{n,s,i} \mathbf{x}_{n,\ell}$, (3) network and non-network regressors of the form $\mathbf{w}_{n,m\lambda,i} \mathbf{x}_{n,q} \mathbf{x}_{n,i,\ell}$, and (4) network regressors and network outcomes of the form $\mathbf{w}_{n,m,\lambda,i} \mathbf{x}_{n,q} \mathbf{w}_{n,s,i} \mathbf{y}_n$. Thus, to proof the convergence result in (B-33), it suffices to show the convergence for each of the four components described before. Firs, from Lemma 5, it follows that the network regressors and network outcomes component converges to its population mean. The other three results follow by appropriately choosing $R_{n,i,j}$, $\mathcal{I}_i$, and $w_{i,j}$ to apply the results from Lemma 4. For (1), choose $R_{n,i} = x_{n,i,q} x_{n,i,\ell}$ for two arbitrary regressors q and ℓ, $\mathcal{I}_i = \varnothing$ so that there is not a second summation over j and $w_i = 1$. For (2) and (3), choose $R_{n,i,j} = x_{q,i} x_{\ell,j}$ for arbitrary regressors q and ℓ. Regarding the set of indexes, for (2), choose $\mathcal{I}_i = \mathcal{P}_n(i, 1, m, \lambda) \times \mathcal{P}_n(i, 1, m)$, and for (3) choose $\mathcal{I}_i = \mathcal{P}_n(i, 1, m, \lambda)$ and corresponding weights if the relevant network is $\mathcal{W}_{n,m,\lambda}$, or $\mathcal{I}_i = \mathcal{P}_n(i, 1, m)$ and corresponding weights if the relevant network is $\mathbf{W}_m$. Therefore, applying Lemma 4 component-wise for $\sum_{i \in \mathcal{I}_n} \mathbf{z}_{n,i} \mathbf{d}_{n,i}^\top$ implies the result. ∎

Lemma 7 (LLN for Instruments and Errors) *Let Assumptions 2, 6, 7 and 8 hold. Then as $n \to \infty$*

$$\left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \mathbf{z}_{n,i} \varepsilon_{n,i}^\top - \mathbb{E}[\mathbf{z}_{N,i} \varepsilon_{N,i}^\top] \right\|_1 \xrightarrow{a.s.} 0. \tag{B-34}$$

Proof. Given that $\mathbf{r}_{n,i} = [\mathbf{x}_i, \varepsilon_i]$ and $\mathbf{z}_i$ can be divided into network and non-network components, the proof of this result is analogous for that of Lemma 6 parts (1) and (3). ■

Lemma 8 (Central Limit Theorem) *Let Assumptions 2, and 6 to 9 hold. Define the sum $S_n = \sum_{i \in \mathcal{I}_n} z_{n,i,q} \varepsilon_{n,i}$, where $z_{n,i,q}$ is the qth entrance of the vector $\mathbf{z}_{n,i}$, $\mathcal{I}_{n,q}$ is the set of individuals with non-zero values in column q of the matrix $\mathbf{Z}_n$. By definition of $\mathbf{z}_i$, $\mathbb{E}[z_{n,i,q}\varepsilon_{n,i}] = 0$. Define $\sigma_n = Var(S_n)$. Then, as $n \to \infty$*

$$\sup_{t \in \mathbf{R}} \left| \mathbf{P}\left\{ \frac{S_n}{\sigma_n} \leq t \mid \mathcal{C}_n \right\} - \Phi(t) \right| \xrightarrow{a.s.} 0,$$

where Φ denotes the distribution function of a $\mathcal{N}(0,1)$,

Proof. Let $Y_{n,i} = z_{n,i,q}\varepsilon_{n,i}$, from Lemma 3, the covariance of any two $Y_{n,i}$ and $Y_{n,j}$ is bounded. Then, the proof follows from applying Lemmas A.2 and A.3 in (Appendix A in Kojevnikov et al., 2020, pp. 899-907) to $Y_{n,i} = z_{n,i,q}$ and S_n/σ_n, respectively. ■

Lemma 9 (Finite Variance) *Define $\mathbf{S}_n = \mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n$ and $\boldsymbol{\Omega}_n = \mathrm{Var}(n^{-1/2}\mathbf{S}_n)$. Let Assumptions 2, and 6 to 8 hold, then $\boldsymbol{\Omega}_n \to \boldsymbol{\Omega}_N < \infty$ as $n \to \infty$.*

Proof. As defined before $n^{-1/2}\mathbf{S}_n = n^{-1/2}\sum_{i=1}^n \mathbf{z}_{n,i}\varepsilon_{n,i}$. The bounded covariance assumptions from Lemma 3 combined with the arguments in Lemma 4 guarantee that $\lim_{n\to\infty} n^{-1}\mathrm{Var}\left(\sum_{i=1}^n \mathbf{z}_{n,i}\varepsilon_{n,i}\right)$ is finite. In particular, from equation (B-27), using the appropriate values for $R_{n,i,j}$, $\mathcal{I}_{n,m,\lambda}$, $\mathcal{I}_i$, $n_{m,\lambda}$, and $w_{i,j}$ (see Lemma 6), it follows that $\mathrm{var}(\sum_{i=1}^n \mathbf{z}_{n,i}\varepsilon_{n,i}) = \mathcal{O}_p(1)$. Given that $\boldsymbol{\Omega}_n$ converges to a finite quantity, it follows that $\boldsymbol{\Omega}_n \to \boldsymbol{\Omega}_N$, where

$$\boldsymbol{\Omega}_N = \lim_{n\to\infty} n^{-1}\left[\sum_{i=1}^n \mathrm{var}(\mathbf{z}_i\varepsilon_i) + \sum_{i\neq j}\mathrm{cov}(\mathbf{z}_i\varepsilon_i, \mathbf{z}_j\varepsilon_j)\right] < \infty.$$

■

Lemma 10 (Multivariate Central Limit Theorem) *Let Assumption 2, and 6 to 9 hold. Then,*

$$n^{-1/2}\sum_{i=1}^n \mathbf{z}_{n,i}\varepsilon_i \xrightarrow{d} \mathcal{N}(0, \Omega_N) \quad as \quad n \to \infty$$

Proof. From Lemma 8, $n^{-1/2}\sum_{i=1}^n z_{n,i,q}\varepsilon_{n,i} \xrightarrow{\mathrm{d}} \mathcal{N}(0,\sigma_n^2)$. From Lemma 9, $\boldsymbol{\Omega}_N$ exists. Therefore, the result follows from the Cramér-Wold device. ■

Appendix C Algorithms to Construct the Moment Condition Matrices

This section describes the computation process required to calculate the empirical analog of the moment condition matrices based on Balasubramanian et al. (2022). First, I use Algorithm 1 to calculate the multilayer shortest paths between a source node s and all the other nodes $v \in \bigcup_{m \in M} V_m$. The time complexity of this algorithm is $\mathcal{O}(V + E \log E)$, where $V \equiv \bigcup_{m \in M} V_m$ and $E \equiv \bigcup_{m \in M} E_m \bigcup \mathcal{C}$. I then use parallel computation to repeat the process for all possible sources. Thus, the described procedure provides all shortest path lengths and edge type changes for the sampled nodes in the multilayer network. With this information, it is feasible to evaluate equation 7 for all possible dyads efficiently.

Algorithm 1: Multilayer Colored Shortest Path

Input: (1) a multilayer graph $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ with non-negative edge weights, and (2) a source vertex $s \in \bigcup_{m \in M} V_m$.

Output: shortest paths and color changes for all nodes $v \neq s$ in $\mathcal{M}$.

Initialization $Q = \bigcup_{m \in M} V_m$, $D = \infty$ for all nodes in Q, $P = CP = CC = \varnothing$ Define $D[s] = 0$;

while *Q is not empty* **do**

- u= node in Q with the minimum distance to s; Remove u from Q and add it to P; **foreach** *v directly connected to u* **do**
  - distance $=$ $D[u]+$ weight of the edge from u to v;
  - **if** *distance $<= D[v]$* **then**
    - $D[v]$ =distance
    - **if** $s = v$ **then**
      - CP=layer where the edge exists and $CC = 0$;
    - **else**
      - CP=$CP[u]$+layer where the edge exists and CC = number of edge-layer changes

---

The calculation of the second shortest path involves a recursive execution of the Multilayer Colored Shortest Path algorithm. The basic idea is that for each node, I replace weighs of the edges in the shortest path for any arbitrary node v for infinite and then run Algorithm 1 again. Algorithm 2 details the process. This algorithm has the same complexity as the one in 1. The construction of the matrices $\mathcal{W}_{m,\beta}$ and $\mathcal{W}_{m,\delta}$ for all $m \in 1, \ldots, M$ is then only a matter of filtering the dyads that fulfill the requirements of equations 6 and 7.

Algorithm 2: Multilayer Colored Second Shortest Path

Input: (1) a multilayer graph $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ with non-negative edge weights, and (2) a source vertex $s \in \bigcup_{m \in M} V_m$.

Output: second shortest paths and color changes for all nodes $v \neq s$ in $\mathcal{M}$.

foreach v $s \in \bigcup_{m \in M} V_m$ **do**
- Multilayer Colored Shortest Path for v;
- store shortest path;
- **foreach** u *in the shortest path* **do**
  - replace the edge weights for ∞;
- **do** Multilayer Colored Shortest Path for v again;

Appendix D Data and Additional Empirical Results

Appendix D.1 Scholars Network Constructions

Co-authors – This type of connection happens when 1) a scholar publishes a paper (alone or with other authors) in any of the four journals under consideration 2) in any year of interest, 3) also publishes other papers (with the same or other co-authors) in any of the four journals and in the seven-year timeframe considered here. Then a co-autorship connection is created between this scholar and all of his/her co-authors in these multiple publications. For example, scholar 5 published a paper in JPE with scholars 424, 436 and 1,041 in 2001. Additionally, the same scholar 5 published another article in the AER in 2005, this time in collaboration with scholar 1,108. Moreover, in addition to the article in JPE with authors 5, 436 and 1,041, scholar 424 wrote another paper in ECA along with authors 847 and 889. Thus, scholar 5 is said to have a co-authorship connection with 424, 436, 1,041, and 1,108, as long as the year of interest is 2005 or 2006 because in those year all the publications obtained during the previous years are also considered. Similarly, scholars 424, 436, and 1,041 are connected between them, and scholar 424 is also connected with 847 and 889, who are also co-authors themselves.

Alumni – This type of connection happens when two scholars got their Ph.D. degrees from the same university and within a maximum of a three-year gap. For example, scholars 1 and 1,699 have an alumni connection equal to one using this criteria because both completed their Ph.D. degrees at Princeton University in 1995.

Ph.D. Advisor – This type of connection happens when a scholar studying in an institution obtains his or her Ph.D. in a particular year when another scholar held the position of assistant, associate, or full professor in the same institution and also shares at least one common JEL code. Of the 2,057 scholars, 46% are full professors, 16% are associate professors, while 23% are assistant professors. 15% of scholars held other positions such as post doc, visiting, etc. For example, scholar 2 was employed by University of California Berkeley as a full professor between 2000-2006, working in Law and Economics

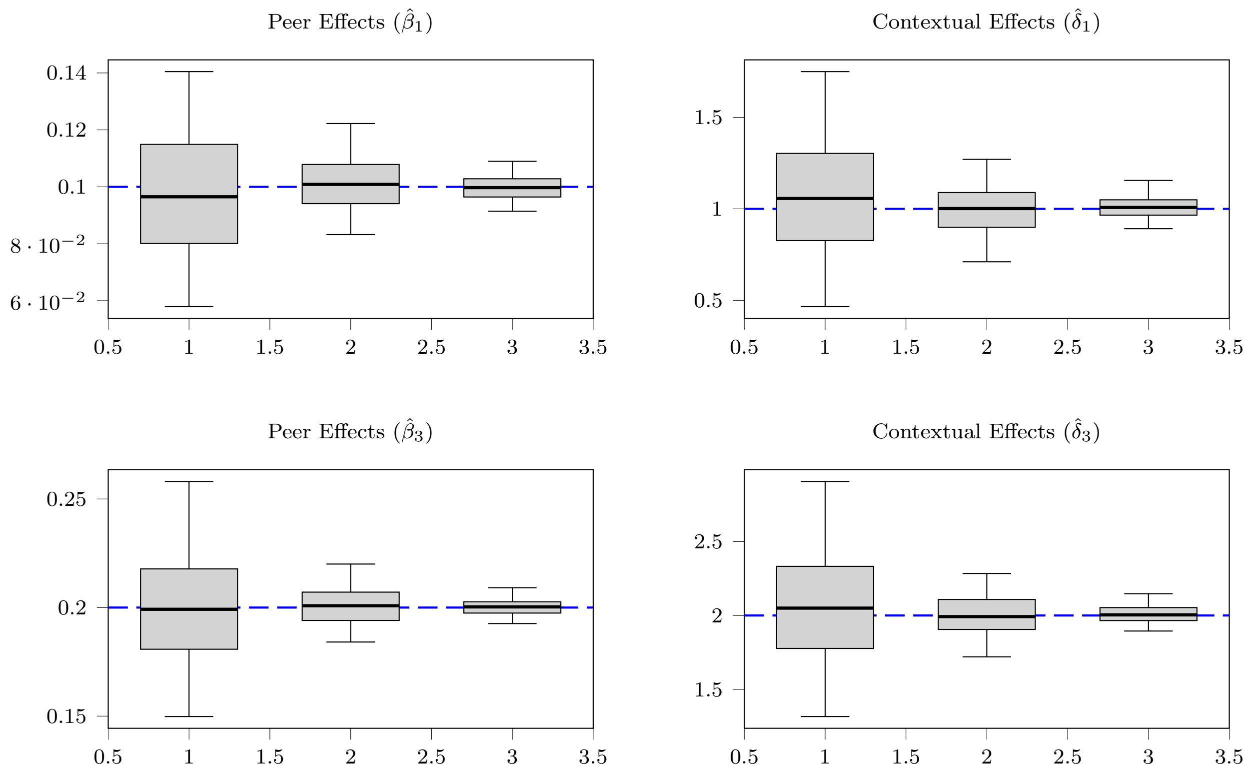
(JEL code: K) and Industrial Organization (JEL code: L). In 2001, author 1,035 finished her Ph.D. in Industrial Organization (JEL code: L) and Public Economics (JEL code: H). Then author 2 and 1,035 are said to have a Ph.D. Advisor connection. There are 19 scholars who held the title Assistant Professor in the same institution they obtained their Ph.D. in the same year. Their Ph.D. Advisor connection with the remaining 2,038 scholars and among themselves are set to zero.

Colleagues – This link happens when two scholars are considered each other's colleague, i.e., they ever worked in the same institution in the same time period. For instance, scholars 1 and 103 are connected because both were working at the University of Illinois Urbana-Champaign between 2000 and 2002. As in *Same Ph.D.*, 136 scholars are omitted from the estimating sample because institution information for them is missing.

Appendix D.2 Additional Simulation and Estimation Results

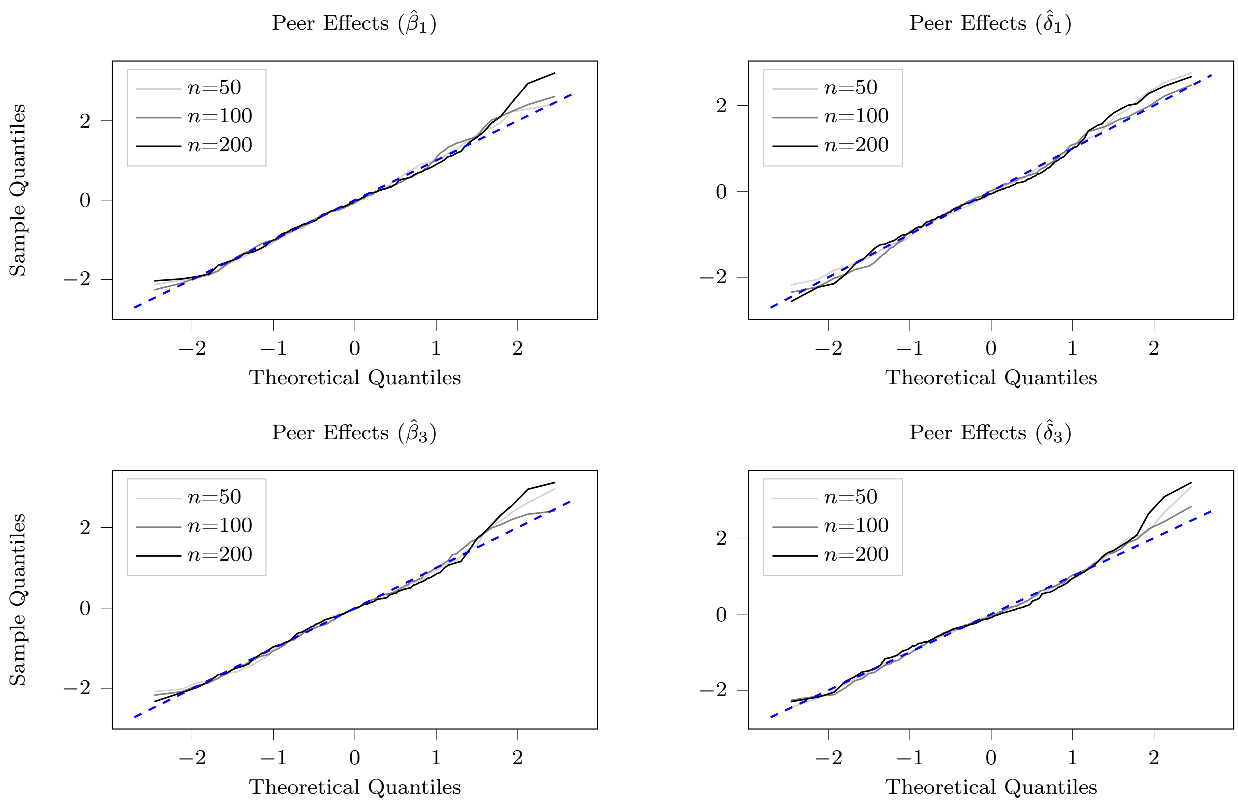
This section contains additional simulation and estimation results. Plots 5 and 6 show the results from the Monte Carlo simulation for the additional network effects $[\beta_1^0, \beta_3^0, \delta_1^0, \delta_3^0]$. The results for the other parameters are presented in Section 6 in the main text. The results here are consistent with those for the parameters $[\beta_2^0, \delta_2^0]$. The simulation confirms the desirable properties of the estimator in finite sample. Similarly, Tables 3 and 4 present the empirical estimation results using the publications data described in section 7. These results are used as a benchmark to compare the behavior of the efficient GMM estimator.

Figure 5: Box Plots for the GMM Estimator



Note: Box plots in the first row depict the Monte Carlo performance of the proposed GMM estimator. The boxplots are based on 1,000 for sample sizes $n \in \{50, 100, 200\}$. The whiskers display the 5% and 95% empirical quantiles.

Figure 6: Q-Q plots of the GMM Estimator



Note: Q-Q plots in the second row are based on the standardized sample of 1,000 Monte Carlo replications of the proposed GMM estimator of the parameters in (1) and sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black). The blue dashed line shows the 45 degree line.

Table 3: GMM Estimation Results for Social and Direct Effects

	2002	2003	2004	2005	2006
Peer Effects ($\{\widehat{\beta}\}_{m=1}^{3}$)					
`Co-authors`	0.485***	0.428***	0.566***	0.536***	0.468*
	(0.123)	(0.112)	(0.114)	(0.106)	(0.279)
`Alumni`	0.120	0.113	0.165	0.175	0.374
	(0.262)	(0.216)	(0.263)	(0.172)	(0.387)
`Advisor`	0.033	0.197	-0.606*	-0.437*	-0.086
	(0.259)	(0.272)	(0.330)	(0.238)	(0.369)
`Colleagues`	0.132	-0.283*	-0.042	0.081	0.086
	(0.471)	(0.1421)	(0.103)	(0.098)	(0.159)
Contextual Effects ($\{\widehat{\boldsymbol{\delta}}\}_{m=1}^{3}$)					
`Co-authors: Editor in Charge`	0.215	-0.288	-0.317	-0.501	0.885
	(0.699)	(0.639)	(0.681)	(0.550)	(0.993)
`Alumni: Editor in Charge`	-0.069	0.069	0.029	0.060	0.402
	(0.216)	(0.156)	(0.138)	(0.128)	(0.316)
`Advisor: Editor in Charge`	0.417	0.105	1.589**	1.071	0.337
	(0.724)	(0.744)	(0.740)	(0.676)	(0.802)
`Colleagues: Editor in Charge`	0.145	0.001	-0.204	-0.293	-0.321
	(0.285)	(0.185)	(0.174)	(0.187)	(0.396)
`Co-authors: Different Gender`	-0.929	-0.753	-1.627	-1.215	2.689
	(1.695)	(1.253)	(1.606)	(0.856)	(1.825)
`Alumni: Different Gender`	-0.385	-0.284	-0.301	-0.014	-0.951
	(0.428)	(0.379)	(0.424)	(0.307)	(0.892)
`Advisor: Different Gender`	1.875	4.991***	5.954***	3.595**	1.382
	(3.116)	(1.734)	(1.884)	(1.515)	(1.869)
`Colleagues: Different Gender`	0.606	0.253	0.015 -0.357	-2.146	
	(0.522)	(0.552)	(0.567)	(0.732)	(1.633)
Contextual Effects ($\widehat{\gamma}$)					
`Editor in Charge`	-0.069	-0.094	-0.042	-0.017	0.061
	(0.135)	(0.115)	(0.139)	(0.123)	(0.141)
`Different Gender`	0.253**	0.229**	0.189**	0.143*	0.069
	(0.122)	(0.102)	(0.091)	(0.081)	(0.093)
`Number of Pages`	0.025***	0.023***	0.020***	0.017***	0.016***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
`Number of Authors`	0.064	0.084*	0.052	0.074*	0.065*
	(0.056)	(0.047)	(0.043)	(0.038)	(0.035)
`Number of References`	0.009***	0.009***	0.007***	0.008***	0.011***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
`Co-authors: Isolated`	1.566***	1.367***	1.775***	1.629***	1.452
	(0.437)	(0.404)	(0.409)	(0.384)	(1.032)
n	729	961	1187	1412	1628

Note: Standard errors are in parenthesis and are calculated using the network HAC estimator of the covariance matrix in equation (5.1) where the function K is the Parzen kernel and the bandwidth $D_n = 1.8 \times [\log(\text{ avg.deg } \vee (1.05))]^{-1} \times \log n$. Stars follow the key: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$, where p stands for p-values. R^2 are calculated as the squared of the sample correlation coefficients between the observed outcomes and their fitted values. All specifications include indicator variables for Journal and Year. The indicator for isolated nodes in the Alumni, Advisor and Colleagues networks are also included but are not statistically significant.

Table 4: OLS Estimation Results for Social and Direct Effects

	2002	2003	2004	2005	2006
Peer Effects ($\{\widehat{\beta}\}_{m=1}^{3}$)					
`Co-authors`	0.359***	0.397***	0.461***	0.452***	0.453***
	(0.061)	(0.049)	(0.047)	(0.045)	(0.042)
`Alumni`	0.026	0.088	0.113*	0.154**	0.139**
	(0.087)	(0.076)	(0.064)	(0.064)	(0.063)
`Advisor`	-0.123*	-0.117*	-0.075	-0.055	-0.045
	(0.07)	(0.066)	(0.061)	(0.054)	(0.056)
`Colleagues`	0.266**	0.096	0.138**	0.156**	0.028
	(0.122)	(0.105)	(0.072)	(0.068)	(0.048)
Contextual Effects ($\{\widehat{\boldsymbol{\delta}}\}_{m=1}^{3}$)					
`Co-authors: Editor in Charge`	-0.219	-0.141	-0.186	0.055	0.086
	(0.215)	(0.202)	(0.178)	(0.180)	(0.172)
`Alumni: Editor in Charge`	-0.063	0.076	-0.012	0.025	0.001
	(0.193)	(0.139)	(0.117)	(0.117)	(0.108)
`Advisor: Editor in Charge`	0.519	0.278	0.202	-0.065	-0.002
	(0.373)	(0.308)	(0.299)	(0.296)	(0.279)
`Colleagues: Editor in Charge`	0.207	0.038	-0.027	-0.092	-0.072
	(0.235)	(0.169)	(0.158)	(0.142)	(0.161)
`Co-authors: Different Gender`	-0.453	-0.265	-0.319	-0.359	-0.048
	(0.421)	(0.388)	(0.368)	(0.387)	(0.432)
`Alumni: Different Gender`	-0.181	-0.162	-0.035	-0.122	-0.214
	(0.27)	(0.229)	(0.219)	(0.204)	(0.211)
`Advisor: Different Gender`	1.434*	2.120***	0.753	0.939	1.290
	(0.786)	(0.742)	(0.971)	(0.974)	(0.964)
`Colleagues: Different Gender`	0.386	0.325	0.012	-0.118	0.112
	(0.405)	(0.393)	(0.379)	(0.974)	(0.964)
Contextual Effects ($\widehat{\boldsymbol{\gamma}}$)					
`Editor in Charge`	-0.017	-0.057	-0.047	0.009	0.044
	(0.126)	(0.112)	(0.131)	(0.114)	(0.107)
`Different Gender`	0.260**	0.206**	0.179	0.133*	0.121*
	(0.113)	(0.094)	(0.082)	(0.078)	(0.069)
`Number of Pages`	0.026***	0.023***	0.019***	0.016	0.016
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
`Number of Authors`	0.054	0.078*	0.071*	0.084**	0.063**
	(0.055)	(0.044)	(0.038)	(0.035)	(0.031)
`Number of References`	0.009***	0.009***	0.009***	0.009***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
`Co-authors: Isolated`	1.129***	1.258***	1.383***	1.328***	1.312***
	(0.236)	(0.196)	(0.183)	(0.178)	(0.164)
n	729	961	1187	1412	1628

Note: Standard errors are in parenthesis and are calculated using the network HAC estimator of the covariance matrix in equation (5.1) where the function K is the Parzen kernel and the bandwidth $D_n = 1.8 \times [\log(\text{ avg.deg } \vee (1.05))]^{-1} \times \log n$. Stars follow the key: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$, where p stands for p-values. R^2 are calculated as the squared of the sample correlation coefficients between the observed outcomes and their fitted values. All specifications include indicator variables for Journal and Year. The indicator for isolated nodes in the Alumni, Advisor and Colleagues networks are also included but are not statistically significant.