Distribution Agreement

Causal Inference in Multilayered Networks

By

Juan Estrada

Doctor of Philosophy

Economics

Kim P. Huynh, Ph.D.

Advisor

David T. Jacho-Chavez, Ph.D.

Advisor

Christoph Breunig, Ph.D.

Committee Member

Kyungmin (Teddy) Kim, Ph.D.

Committee Member

Elena Pesavento, Ph.D.

Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D.

Dean of the James T. Laney School of Graduate Studies

Date

Causal Inference in Multilayered Networks

By

Juan Estrada

M.Sc. in Economics, Emory University, 2020

Advisor:
Kim P. Huynh, Ph.D.
David T. Jacho-Chavez, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
2022

Abstract

Causal Inference in Multilayered Networks

By Juan Estrada

Social and professional networks are critical to determining agents' choices in various contexts, ranging from innovation and trade to labor markets and educational achievement. The existence of network influences on decision-making processes can affect aggregate outcomes and the effects of policy intervention. Therefore, understanding whether network structures affect individuals' outcomes is relevant in social sciences such as economics, public health, sociology, and political science. However, empirically testing the existence of network effects with observational data becomes a challenge because of outstanding identification issues such as endogenous network formation. This dissertation provides novel methodologies to causally estimate network effects with observational data robust to network endogeneity issues. Differing from existing approaches, the methods I propose are semiparametric, do not require to specify a structural network formation model, and allow for the identification and estimation of heterogeneous network effects generated by different types of social/professional links.

The first chapter proposes a method to identify and estimate the linear model of peer effects parameters when a predetermined set of exogenous connections induce the observed interest network. The second chapter expands the results from the first chapter by allowing the possibility of multiple types of potentially endogenous networks to affect individual outcomes. I show that the identification of heterogeneous network effects is possible under the assumption that the dependence between individuals in the population vanishes with their distance in the network space. Finally, chapter three focuses directly on the process of determining the formation of network structures rather than their effect on outcomes. In particular, it provides a novel approach to identify and perform inference on the utility parameters of a network formation model with payoff externalities using observed network data.

Causal Inference in Multilayered Networks

By

Juan Estrada

M.Sc. in Economics, Emory University, 2020

Advisor:
Kim P. Huynh, Ph.D.
David T. Jacho-Chavez, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
2022

Acknowledgments

I am successfully reaching the final phase of my Ph.D. This achievement was possible thanks to the help and support of an outstanding group of people. Mentors, professors, friends, and family have been essential to my formation as a researcher, and I want to express my gratitude. I want to dedicate this dissertation, the product of my efforts during the last five years, to the memory of my father, Juan Rafael Estrada. Thanks for teaching me to be curious and love knowledge; those lessons will always be with me.

To David, my mentor, you keep pushing me to improve myself as a researcher and a professional. Thank you for always being there for advice. I could not have asked for a more diligent and hard-working advisor. Thank you, Kim, for your support and for exposing me to the economics profession as much as you did. You two have been an excellent source to learn how to be an effective researcher. Thank you, Elena, for all your advice and immense willingness to help. I appreciate that you are keen to reach out to others to keep helping me advance my career as a researcher. Thanks to Christoph and Teddy for their comments and suggestions. Thanks to Cliff for his confidence and enthusiasm with our new projects.

Thank you, Renee, Stephanie, Marie, Felecia, Elizabeth, and the Economics department staff for your excellent assistance with my many administrative requests. I want to acknowledge the impact that academic conferences have had on the development of my research. The 2019 Advances in Econometrics workshop on the Econometrics of Networks co-hosted by the Research Centre of the Lucian Blaga University of Sibiu and the National Bank of Romania. Thanks to the Co-editors and organizers of the conference: Áureo de Paula (University College London), Elie Tamer (Harvard University), Marcel-Cristian Voia (Laboratoire d'Économie d'Orléans) for providing early feedback on my research. The 2019 Microeconometrics: Survey Methodology and Data Science conference organized by the Bank of Canada (BoC) in Ottawa from 26-27 September 2019. Thanks to Heng Chen and Kim P. Huynh for organizing an excellent conference that brought together a diverse set of researchers, including Harry Crane, which influenced my thinking of multilayer networks. 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. In particular, I thank Bruno Ferman, Guido Kuersteiner, Alejandro Sanchez Becerra, Pietro Spini, and Davide Viviano for critical comments that helped me improve my first two chapters.

Thank you to the members of the econometrics of networks study group, Pablo, Alexandra, Santiago, and Diego. Those discussions and conversations have been vital in developing my research ideas. Special thanks to Pablo, Santiago, and Diego, who have been my close friends in this process. You are a positive source of peer effects. To my dear friend and colleague Cheng, we have gone through a lot together, and I appreciate having you by my side. We will keep our work going. Thanks for keeping me up to date with my econometrics knowledge to all my junior fellow graduate students. It has been a pleasure to share the review session time with you.

I want to thank the computational advice and support provided by Kim P. Huynh, Nathalie Swift, and the Bank of Canada High-Performance Computing team. In particular, I want to acknowledge Venkataraman Balasubramanian for providing his invaluable expertise for Chapters One and Two of my dissertation. In Chapter One, for implementing and optimizing the Python code for the high-dimensional matrix computations. In Chapter Two, for creating and developing the C++ for the shortest path algorithm. This efficient implementation made computation of the multilayer network feasible and efficient. Venkat also spent considerable development testing to tune the code to the Bank of Canada's EDITH2 High-Performance Cluster. I hope that we can continue to collaborate and publish the algorithmic research on multilayer networks. The views expressed in this dissertation are solely mine and may differ from official Bank of Canada views.

Finally, I want to express my appreciation for my family and friends' support through this process. To my mother and sister, Soledad, and Isabel, you have always been there for me, showing me the meaning of strength and resilience. I am here because you continuously reminded me of the value of pursuing my goals. To Kaylyn, you are part of my family. Thanks for always being the best motivation to keep me going. You taught me how to be a better professional, a better presenter, and a better writer. You have been a continuous source of inspiration. To Kevin and Marlene, you have been my family in the US and taught me to love this country. To my friends in Colombia and Europe, Felipe, Pablo, Mateo, Santiago Rodriguez, Santiago Vasquez, and Simon, you are a happy place where I can always go back.

Contents

List of Figures

List of Tables

Chapter 1

Instrumental Network Estimation of Social Effects

Note: This chapter is the product of collaborative research with Kim P. Huynh, David T. Jacho-Chávez and Leonardo Sánchez-Aragón. The views expressed in this chapter are solely those of the authors and may differ from official Bank of Canada views.

This paper proposes a new method to identify and estimate the parameters of linear models of peer effects in situations where predetermined exogenous peers induce the observed network of interest. We argue that exogenous predetermined networks do not generate social effects. However, based on the initial set of connections, agents can endogenously form relevant relationships that can create peer influences. In this context, we introduce a moment condition that aggregates local heterogeneous identifying information for all the individuals in the population. We show that it is possible to identify the parameters of interest by using the exogenous variation in the predetermined groups and provide a consistent GMM estimator *and root-n asymptotic normal. We also characterize the asymptotically efficient variance-covariance matrix that considers the network dependence among individuals. Monte Carlo exercises confirm the adequate finite sample performance of the proposed estimator, and an empirical application finds positive and significant peer effects in citations among research articles published in top general interest journals in economics.*

1.1 Introduction

Network structures shape observed outcomes across different types of markets. Empirical evidence shows that connected individuals tend to have similar outcomes, see e.g., Sacerdote (2001a) in economics, Alexander et al. (2020) in Public Health, Kreager et al. (2020) in Criminology, and Salmivalli (2020) in Sociology to mention just a few. The observation that connected individuals have correlated outcomes is consistent with models where individuals decision-making process depends on the behavior of their connections. Blume et al. (2015) present a model where individual utility is a function of both the peer characteristics –*contextual effects*– and peer actions or outcomes –*peer effects*. The workhorse model in the Social Sciences for this type of setting is the so-called *linear-in-means* regression model proposed by Manski (1993), (see Section 3.1 in de Paula, 2017, pp. 275-289) where the outcome variable for observation i (e.g., a person, a firm, or a country), y_i, is determined according to

$$y_i = \alpha + \beta \sum_{r=1}^{n} \mathbf{W}_{ir} y_r + \delta \sum_{r=1}^{n} \mathbf{W}_{ir} \mathbf{x}_r + \gamma \mathbf{x}_i + v_i \tag{1.1}$$

where $i, j \in \{1, \ldots, n\}$ are also known as *nodes*, $\mathbf{x}$ is a vector of attributes that characterizes the observations i and j, $w_{i,j} = 1$ if j is connected with i (an *edge*), and 0 otherwise, v_i represents an unobserved latent error, and n is the number of observations or nodes in the sample. The structure of the *social* network is fully characterized by the square $n \times n$ matrix, $\mathbf{W}$, with (i, j) entry given by $w_{i,j}$, i.e., the adjacency matrix. The structural parameters β and $\boldsymbol{\delta}$ capture the peer and contextual effects respectively, while $\boldsymbol{\gamma}$ captures the direct effects of the observation's own characteristics. They are jointly known as the social (or neighbor) effect parameters and the object of interest in empirical studies with network data.

Sufficient conditions under which the social effects parameters in (1.1) can be uniquely recovered from the estimating sample $\left\{y_i, \mathbf{x}_i^\top, \{w_{i,j}\}_{j=1,j\neq i}^n\right\}_{i=1}^n$ is well understood under the assumption that the adjacency matrix, $\mathbf{W}$, is exogenous (de Paula, 2017). However, the endogenous case remains an active area of research among social scientists because of simultaneity bias, measurement error in the link information, or because there might be a natural correlation between covariates and the error term in (1.1) (Johnsson and Moon, 2019).

This research provides a new way to estimate social effects from observational data consisting of characteristics and connections among individual-level observations in contexts in which individuals freely form the social relationships of interest after being exogenously assigned to peer groups, see, e.g., Carrell et al. (2009), Carrell et al. (2011), and Carrell et al. (2013). The proposed estimator of the social effects turns out to be a linear Generalized-Method-of-Moments (GMM) estimator that has an explicit formula, converges at the standard rate, and has a multivariate normal limiting distribution with an efficient asymptotic variance-covariance matrix that takes into account the network dependence in the data. Specifically, I proposed to conduct inference using the results in Kojevnikov et al. (2020)–that functionals of network data become uncorrelated as observations becomes distant in the network space; i.e., ψ–dependence. By imposing ψ–dependence ob the data-generating process, I propose a Heteroskedasticity and Autocorrelation Consistent (HAC) estimator of the efficient asymptotic variance-covariance matrix that corrects for the presence of network dependence in the data.

Recent articles offering methods to alleviate the network endogeneity issue propose augmenting 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), Johnsson and Moon (2019) and Auerbach (2022). Specifying an additional network formation model generally involves imposing strong functional form and parametric assumptions (Graham and Pelican, 2020). This paper contributes by providing a computationally-simple estimator that does not require imposing extra structure to the model, which at the same time takes into account the potential resorting process after the initial exogenous configuration.

I provide a Monte Carlo exercise that showcases the versatility of the proposed estimator by analyzing its performance in two separate data generating processes that consider the unobserved homophily in the network formation and measurement error scenarios. The simulation results show that the proposed GMM estimator performs well in terms of bias and asymptotic normal approximation, with sample sizes as small as 50 observations. In addition to the simulated data, this paper also presents an application to real data on publication outcomes in Economics. The use of web scraping and existing data on authors' research fields, education, and employment history allows the creation of two types of professional ties among scholars, namely co-authorship and alumni connections. These multiplex networks are then used to uncover positive and significant peer effects in terms of citations among articles published by these scholars, 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 after controlling for other articles' characteristics such as number of pages, number of bibliographic references, co-authoring with current and previous editors, and various network fixed effects.

The remainder of this article is organized as follows. Section 1.2 introduces the necessary definitions, including distance metrics on graph space, as well as the general notation used throughout the paper. Section 1.3 then provides sufficient conditions under which social parameters can be uniquely recovered from observational data when information from an exogenous network is available. The proposed estimator and its asymptotic distributional properties under ψ–dependence are presented in Section 1.4. A simulation study is described in Section 1.5, while Section 1.6 provides an empirical illustration of the methods discussed by applying them to a data set of publication outcomes in economics. Section 1.7 concludes with a discussion on various venues the proposed method can be used for experimental designs for example.

1.2 Preliminaries

Full observability of two types of network is assumed here. One is the endogenous network of interest that can create social externalities, and the other is the instrumental network induced by random assignment. We use a *multilayered* network data structure to characterize our data generating process, see, i.e., Boccaletti et al. (2014) and Kivela et al. (2014) for up-to-date comprehensive surveys and references therein. Following Boccaletti et al.'s (2014) definition, a multilayer network is a pair $\mathcal{M} = (\mathcal{N}, \mathcal{C})$, where $\mathcal{N} = \{\mathcal{G}_m; \quad m \in \{1, \dots, M\}\}$

is a family of M graphs and $\mathcal{C}$ is the set of interconnections between nodes of different layers g_α and g_β with $\alpha \neq \beta$. When the same nodes are in each layer, and there are no connections between different nodes in different layers except with themselves, these networks are called *multiplex*, see, i.e., Atkisson et al. (2020). In our case we have $m = 2$, and we denote $\mathcal{G}_N$ to be the population network of interest, and $\mathcal{G}_{N,0}$ the population instrumental network, where N is the number of individuals in the population, and it is assumed to be arbitrarily large. Similarly, as in Graham (2020), the observed networks of size $n < N$ are denoted as $\mathcal{G}_n$ respectively, and are assumed to coincide with the subgraphs induced by a sample of nodes from their corresponding large population networks.

In the population, we represent the two layers $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$ in the multiplex network with the adjacency matrices of each layer, i.e., $\mathbf{W}_N = [w_{N;i,j}]$ and $\mathbf{W}_{N,0} = [w_{N,0;i,j}]$, where $w_{N;i,j}, w_{N,0;i,j} \in (0,1]$ are weights representing the importance of the (i,j) connection in each of the networks, and $w_{N;i,j} = 0$ if i and j are not connected in $\mathcal{G}_N$, and similarly for $w_{N,0;i,j} = 0$ in $\mathcal{G}_{N,0}$. We also define the vectors $\mathbf{w}_{N,i} = [w_{N;i,1}, \ldots, w_{N;i,N}]^\top \in [0,1]^N$ and $\mathbf{w}_{N,0,i} = [w_{N,0;i,1}, \ldots, w_{N,0;i,N}]^\top \in [0,1]^N$ to be the ith row of the adjacency matrices $\mathbf{W}_N$ and $\mathbf{W}_{N,0}$, respectively. Adjacency matrices from the observed sample, $\mathbf{W}_n$ and $\mathbf{W}_{n,0}$, are defined and formed accordingly.

Section 2.5 uses the concept of $\psi-$dependence to bound the dependence among individuals as a function of their distance in the network space. Following the literature on graph theory, we use the shortest path length as our measure of distance, i.e., let $d_n(i,j)$ be the minimum path length connecting individuals i and j in the network $\mathcal{G}_n$ induced by the sample of size n. We define the following group of sets based on the geodesic distance $d_n(i,j)$.

Let A and B be any two sets of individuals of sizes $a, b \in \mathbb{N}_+$. Define the distance between sets as $d_n(A, B) = \min_{i \in A} \min_{j \in B} d_n(i, j)$. Consider the following sets: (i) $\mathcal{P}_n^+(a, b, d) = \{(A, B) : A, B \subset \mathcal{I}_n, |A| = a, |B| = b, \text{ and } d_n(A, B) \geq d\}$ containing groups of nodes at distance of at least d from each other, (ii) $\mathcal{P}_n^-(a, b, d) = \{(A, B) : A, B \subset \mathcal{I}_n, |A| = a, |B| = b, \text{ and } d_n(A, B) \leq d\}$ containing groups of nodes at distance of at most d from each other, (iii) $\mathcal{P}_n(a, b, d) = \{(A, B) : A, B \subset \mathcal{I}_n, |A| = a, |B| = b, \text{ and } d_n(A, B) = d\}$ 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 $\mathcal{P}_n^+(i, d) = \{j \in \mathcal{I}_n : d_n(i, j) \geq d\}$, $\mathcal{P}_n(i, d) = \{j \in \mathcal{I}_n : d_n = d\}$ and $\mathcal{P}_n^-(i, d) = \{j \in \mathcal{I}_n : d_n \leq d\}$.

1.3 Peer Effects Model and Identification

The model in the population is formed by a set $\mathcal{I}_N$ of N agents, where $N \subset \mathbb{N}^+$ can be arbitrarily large. In the population, each agent, i, is characterized by a set of K observable characteristics $\mathbf{x}_{N,i}$, and an unobserved idiosyncratic shock (error) $\varepsilon_{N,i}$. Agents in the population are connected by two types of networks $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$. Motivated by the empirical application, we assume that the network $\mathcal{G}_0$ is determined at random by an independent institution that is trying to maximize neither any agent's nor any collective welfare function. However, the network $\mathcal{G}_N$ is formed endogenously by agents who are deciding whether to connect or not with each other to maximize some gain/loss function that potentially depends on others' observed and unobserved characteristics.

We assume that there is a population joint distribution determining the dependence patterns between the regressors, the networks, and the errors denoted by $\mathcal{F}(\mathbf{X}_N, \mathcal{G}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N)$, where $\mathbf{X}_N = [\mathbf{x}_{N,1}, \cdots, \mathbf{x}_{N,N}]^\top \in \mathscr{X}$ and $\boldsymbol{\varepsilon}_N = [\varepsilon_{N,1}, \cdots, \varepsilon_{N,N}]^\top \in \mathbb{R}^N$ represent the matrix of regressors and the vector of errors for all agents in $\mathcal{I}_N$. The joint distribution is characterized by three main features. First, by the properties of randomization, the network $\mathcal{G}_{N,0}$ is independent to the agents' observed and unobserved characteristics $\mathbf{x}_{N,i} \in \mathbb{R}^K$, and $\varepsilon_{N,i} \in \mathbb{R}$. Second, agents randomly assign to the same group in $\mathcal{G}_{N,0}$ are more likely to form connections in $\mathcal{G}_N$, see, e.g., Goldsmith-Pinkham and Imbens (2013). Finally, in general, there could be agents in the population who do not have any connections in one or both networks. We call them *isolated* agents and it is shown here they provide identifying information for both direct as well as social effects.

To incorporate the possibility that individual i can be isolated in one or both of the networks, let $\eta_{N,i}$ and $\eta_{N,0,i}$ be two Bernoulli random variables that equal one if individual i is non-isolated in the networks $\mathcal{G}_N$ or $\mathcal{G}_{N,0}$, respectively. It follows that $\kappa_{N,i} = \mathbb{E}[\eta_{N,i}]$ and $\kappa_{N,0,i} = \mathbb{E}[\eta_{N,0,i}]$ give the unconditional probability that individual i is non-isolated in the respective network, where the expectations are taken with respect to the marginal distributions of $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$, respectively. With this notation, the following assumption imposes restrictions on the joint distribution $\mathcal{F}$ that are consistent with the intuition in our empirical application.

Assumption 1 (Joint Distribution Characterization). *Consider the sets $\mathscr{G}$, $\mathscr{G}_0$, $\mathscr{X}$ of all possible realizations of $\mathcal{G}_N$, $\mathcal{G}_{0,N}$ and $\mathbf{X}_N$ with positive probability mass in $\mathcal{F}$, respectively. The following is true:*

(i) The distribution $\mathcal{H}(\mathbf{X}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N) = \int_{\mathcal{G}_N \in \mathscr{G}} \mathcal{F}(\mathbf{X}_N, \mathcal{G}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N) d\mathcal{G}_N$ is such that $\mathbb{E}[\mathbf{X}_N^\top \boldsymbol{\varepsilon}_N] = \mathbf{0}_K$, where $\mathbf{0}_K$ is a $K \times 1$ vector of zeros.

(ii) $\forall \mathcal{G}_{N,0} \in \mathscr{G}_0$ *and* $\mathbf{X}_N \in \mathscr{X}$*, the conditional probability* $\mathcal{F}(\mathcal{G}_N, \boldsymbol{\varepsilon}_N \mid \mathcal{G}_{N,0}, \mathbf{X}_N)$ *is such that* $\Pr(w_{N;i,j} > 0 | \mathcal{G}_{N,0}, \mathbf{X}_N) = \rho(w_{N,0;i,j}, \mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N)$*, for some real-valued function* $\rho : \mathscr{G}_0 \times \mathscr{X} \times \mathbb{R}^N \to [0,1]$*. Moreover,* $\rho(0, \mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N) < \rho(\mathbb{1}\{w_{N,0;i,j} > 0\}, \mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N)$*,* $\forall (\mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N) \in \mathscr{G}_0 \times \mathscr{X} \times \mathbb{R}^N$*, where* $\mathbb{1}(\cdot)$ *is the usual indicator function that equals one if its argument is true and zero otherwise.*

(iii) $\forall \mathcal{G}_{N,0} \in \mathscr{G}_0$ *and associated marginal distributions, the random variables* $\{\eta_{N,0,i}\}_{i=1}^N$ *are such that the event* $\eta_{N,0,i} = 0$*,* $\forall i \in \mathcal{I}_N$ *happens with probability zero. (iv)* $\forall \mathcal{G}_N \in \mathscr{G}$ *and associated marginal distributions, the random variables* $\{\eta_{N,i}\}_{i=1}^N$ *are such that the event* $\eta_{N,i} = 0$*,* $\forall i \in \mathcal{I}_N$ *happens with probability zero.*

Assumption $1(i)$ does not impose any restrictions on the correlation among $\mathcal{G}_N$, $\mathbf{X}_N$, and $\boldsymbol{\varepsilon}_N$. In particular, this assumption allows for an endogenous network formation process where agents can create connections in $\mathcal{G}_N$ based on observed and unobserved characteristics, which could also induce correlation between $\mathbf{X}_N$ and $\boldsymbol{\varepsilon}_N$ because of homophily. Importantly, the random assignment defining $\mathbf{W}_{N,0}$ guarantees that the network is independent of the regressors and the errors, and conditioning on it is not necessary. Assumption $1(ii)$ imposes the condition that the probability of agents i and j being connected in $\mathcal{G}_N$ increases when they are connected in $\mathcal{G}_{N,0}$. However, we are agnostic about the dependence structure in the implicit network formation process for which we do not impose any explicit functional form. In particular, this assumption can accommodate the pairwise independent network formation models as in Graham (2017), and the network formation models with strategic interactions as in de Paula et al. (2018) and Graham and Pelican (2020). Assumptions $1(iii)$ and $1(iv)$ are necessary conditions for the model to be identified, and exclude the possibility that $\mathbf{W}_{N,0} = \mathbf{W}_N = \mathbf{O}_N$ (the $N \times N$ matrix of zeroes).

There has been a relevant number of articles in economics showing that agents who interact together in social networks tend to exhibit correlated behavior (Sacerdote, 2014). The linear model of peer effects from Manski (1993) has been for many years now the most widely used tool to estimate peer effects in applied work (de Paula, 2017). Manski's linear model and additional variations allowing for network data have a structural interpretation as the best response function of a Bayesian game of social interactions as shown by Blume et al. (2015). In this paper, we take the linear model of peer effects as our primitive model of interest, and we investigate identification under endogenous network formation. We assume that only the connections that are formed optimally by agents can create peer effects.

Assumption 2 (Linear Model). *The optimal choice (outcome), $y_{N,i}$, for agent i is characterized by*

$$y_{N,i} = \beta_0 \sum_{j \neq i} w_{N,i,j} y_{N,j} + \sum_{j \neq i} w_{N,i,j} \mathbf{x}_{N,j}^\top \boldsymbol{\delta}_0 + \widetilde{\mathbf{x}}_{N,i}^\top \boldsymbol{\gamma}_0 + \varepsilon_{N,i}, \tag{1.2}$$

where $\widetilde{\mathbf{x}}_{N,i} = [1, \mathbf{x}_{N,i}^\top]^\top$, $\boldsymbol{\theta}_0 = (\beta_0, \boldsymbol{\delta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$ belongs to the interior of the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^{2K+2}$ which is assumed to be compact.

Assumption 2 is a linear model for social effects, and it effectively imposes an exclusion restriction on the network $\mathcal{G}_{N,0}$. The intuitive motivation for this assumption is that agents would tend to behave similar to others with whom they have a close social relationship. In particular, we argue that a group of people randomly assigned together into a group are not likely to generate peer effects to each other. After the randomization, agents can form endogenously connections that are relevant to create social effects. This intuition has found support in the literature, see, e.g., Carrell et al. (2013) who found that groups designed optimally to improve academic performance ended up having a negative effect because of the role of endogenous link formation. The coefficients β_0 and $\boldsymbol{\delta}_0$ are known as the peer effects and the contextual effects respectively and referred jointly as the social effects parameters hereafter. The assumption that the parameters $\boldsymbol{\theta}_0$ are in the interior of the parameter space is particularly relevant for the coefficient β_0, because (1.2) has a solution in terms of $w_{N,i}$, $\mathbf{X}_N$, and $\varepsilon_{N,i}$ only when $\beta_0 < 1/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{W}_N$. Assuming $K = 1$ and no constant for the sake of illustration, Assumption 2 implies that the peer effects regressor can be written as

$$\mathbf{W}_N\mathbf{y}_N = \gamma_0\mathbf{W}_N\mathbf{x}_N + (\gamma_0\beta_0 + \delta_0)\sum_{p=0}^{\infty}\beta_0^k\mathbf{W}_N^{p+2}\mathbf{x}_N + \sum_{p=0}^{\infty}\beta_0^p\varepsilon_N, \tag{1.3}$$

which under the condition that $\gamma_0\beta_0+\delta_0 \neq 0$ shows that in principle, powers of the adjacency matrix $\mathbf{W}_N$ could be used to instrument $\mathbf{W}_N\mathbf{y}$ (Bramoullé et al., 2009). In the case at hand this approach is not possible because the network $\mathcal{G}_N$ can be endogenous. However, note that from Assumptions 1 (*i*) and (*ii*), powers of the adjacency matrix $\mathbf{W}_{N,0}$ are natural candidates to replace $\mathbf{W}_N$.

We propose to use the random assignment embodied in $\mathbf{W}_{N,0}$ to identify the parameters of a linear model defined on the network space spanned by $\mathbf{W}_N$ in Assumption 2. To formalize our identification strategy, we define $\mathbf{D}_N = [\mathbf{W}_N\mathbf{y}_N, \mathbf{W}_N\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$ to be the matrix of regressors in the matrix-notation counterpart of equation (1.2), and $\mathbf{Z}_N = [\mathbf{W}_{N,0}^p\mathbf{X}_N, \mathbf{W}_{N,0}^{p-1}\mathbf{X}_N, \ldots, \mathbf{W}_{N,0}\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$ to be the matrix producing the moment conditions formed based on Assumptions 1 (*i*) and (*ii*), where $p > 1$ is a constant parameters representing the power of the adjacency matrix used as an instruments. This framework allows for the option of using so called *friends of friends*' characteristics as instruments by letting $p = 2$. The flexibility of Assumption 1 also allows for the use of more indirect connections' characteristics that are at distance $p > 2$. To take into account the fact that individuals without connections at distance p do not provide identifying information for the network effects parameters, define $\eta^p_{N,0,i}$ to be a random variable that equals one if individual i has at least one connection at distance p and zero otherwise. As before, $\kappa^p_{N,0,i} = \mathbb{E}[\eta^p_{N,0,i}]$ is the unconditional probability that individual i has at least one connection at distance p.

Define $\boldsymbol{H}_{N,0,i} = \text{diag}(\eta^p_{N,0,i}, \ldots, \eta_{N,0,i}, 1, \ldots, 1)$ to be a $((p+1)K+1) \times ((p+1)K+1)$ matrix where the first K elements contain the random variables determining whether or not individual i has at least one connection at distance p, the second K elements the random variables determining whether or not individual i has at least one connection at distance $p-1$, and so on, until the last K elements associated with $\mathbf{W}_{N,0}\mathbf{X}_N$, where $\eta_{N,i}$ is the random variable determining whether i is isolated in the network $\mathcal{G}_{N,0}$. Finally, the last $K+1$ elements in the lower-right sub-matrix, which coincide with the non-network regressors $\widetilde{\mathbf{x}}_{N,i}$, are ones. Define the $((p+1)K+1) \times ((p+1)K+1)$ matrix $\boldsymbol{K}_{N,0,i} = \text{diag}(\kappa^p_{N,0,i}, \ldots, \kappa_{N,0,i}, 1, \ldots, 1)$ to be $\mathbb{E}[\boldsymbol{H}_{N,0,i}]$.

From Assumption $1(i)$, it follows that $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}] = 0$ for all i, where $\mathbf{z}_{N,i}$ be the ith row of the matrix $\mathbf{Z}_N$. Defining the function $\boldsymbol{m}_N(\boldsymbol{\theta}) := \sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\varepsilon_{N,i}$, it follows that $\mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta}_0)] = \mathbf{0}$. The following remark about the moment condition $\mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta}_0)]$ is noteworthy. By the law of total expectation $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}] = \boldsymbol{K}_{N,0,i}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i} \mid \boldsymbol{H}^*_{N,0,i} \neq \boldsymbol{O}_{pK}] + (\mathbf{I}_{(p+1)K+1} - \boldsymbol{K}_{N,0,i})\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i} \mid \boldsymbol{H}^*_{N,0,i} = \boldsymbol{O}_{pK}]$, where $\boldsymbol{H}^*_{N,0,i}$ contains the the left top $(pK \times pK)$ upper matrix of $\boldsymbol{H}_{N,0,i}$, and $\boldsymbol{O}_{pK}$ is the $pK \times pK$ zero metrix. Note that when $\boldsymbol{H}^*_{N,0,i} = \boldsymbol{O}_{pK}$, the conditional expectation $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i} \mid \boldsymbol{H}_{N,0,i} = \boldsymbol{O}_{pK}]$ is trivially zero for the first pK elements, and one for that the last $k+1$ positions. Moreover, $\mathbf{I}_{(p+1)K+1} - \boldsymbol{H}_{N,0,i}$ equal zero in the last $k+1$ elements of its diagonal. Therefore, it follows that $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}] = \boldsymbol{K}_{N,0,i}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i} \mid \boldsymbol{H}_{N,0,i} \neq \boldsymbol{O}]$, and thus $\mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta}_0)] = \sum_{i \in \mathcal{I}_N} \boldsymbol{K}_{N,0,i}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i} \mid \boldsymbol{H}_{N,0,i} \neq \boldsymbol{O}]$. This remark is relevant because it gives an interpretation to the unconditional moment condition as a weighted sum where the weights are the probabilities of having distance-p connections and not being isolated. This scheme accommodates heterogeneity on the identifying power of each individual in the population, and gives more importance for agents who are unlikely to not have distance-p connections or to be isolated in the randomized network in the population.

The use of the two networks for identification, requires a relevance condition that guarantees that the two networks have enough overlap. The moment correlating regressors and exogenous variation can be written as $\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_i \mathbf{d}_i^\top]$, where $\mathbf{d}_i$ and $\mathbf{z}_i$ represent the i rows of matrices $\mathbf{D}_N$ and $\mathbf{Z}_N$ respectively. Importantly, we do not assume that $\mathbb{E}[\mathbf{z}_i \mathbf{d}_i^\top]$ are equal for all i. The following Assumption imposes a rank condition related with the strength of the correlation between $\mathbf{z}_i$ and $\mathbf{d}_i$. The conditions in Assumptions 1(*ii*) and 1(*iii*) are necessary but not sufficient for the following assumption to hold:

Assumption 3 (Relevance)**.** *The matrix* $\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top] < \infty$ *has full column rank.*

Assumption 3 imposes restrictions on the product matrices $\mathbf{W}_{N,0}\mathbf{W}_N$ and $\mathbf{W}^p_{N,0}$. As mentioned before, if the product $\mathbf{W}_{N,0}\mathbf{W}_N = \mathbf{O}_N$, identification breaks down. This condition has three important empirical consequences: (1) there should exist nodes in the system that share connections in common from the two networks $\mathcal{G}_{N,0}$ and $\mathcal{G}_N$, (2) it should be possible to connect two nodes with a path formed by an edge from network $\mathcal{G}_{N,0}$ followed by one from the network $\mathcal{G}_N$ (or vise-versa), and (3) the matrices $\mathbf{I}_N$, $\mathbf{W}_{N,0}, \ldots, \mathbf{W}^p_{N,0}$ have to be linearly independent (the network cannot be composed by fully connected groups of the same size). Under the conditions stated before, the following Theorem shows that identification of the parameters of interest in (1.2) is possible, even if the network $\mathbf{W}_N$ is endogenous.

Theorem 1 (Identification)**.** *Let Assumptions 1, 2, and 3 hold, then* $\mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta})] = \mathbf{0}_K$ *if and only if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*, where* $\boldsymbol{m}_N(\boldsymbol{\theta}) := \sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i}\varepsilon_{N,i}$.

1.8 presents the proof for Theorem 1. This Theorem shows that identification is possible in a context where the network of interest $\mathcal{G}_N$ is formed endogenously, by taking advantage of the randomization and exclusion restrictions on the network $\mathcal{G}_{N,0}$. This approach allows to attach causal interpretation to the estimated parameters of a linear model of peer effects that uses observational network data that emerges after an initial randomization. This method can be used to address research designs with randomized peers such as that in Carrell et al. (2013).

1.4 Estimation

We propose a GMM estimator based on the identifying moment condition in Theorem 1. We assume that the analyst observes a sample of size $n < N$ from the population described in the previous section. In our sample scheme, n agents are chosen at random without replacement, and their observed characteristics, outcome, and connections in $\mathcal{G}$ and $\mathcal{G}_0$ are recorded. Therefore, the random sample consist on the observations $\{y_i, \mathbf{x}_i^\top, \{w_{i,j}, w_{0,i,j}\}_{j=1, j\neq i}^n\}_{i=1}^n$, from where it is possible to calculate the $n \times (2K+2)$ matrix of regressors $\mathbf{D}_n$, and the $n \times (3K+1)$ matrix of *instruments* $\mathbf{Z}_n$ (depending on the value of K, the system can be *just* or *over* identified). The population GMM objective function is given by $J_N(\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta})]^\top \mathbf{A}_N \mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta})]$, where $\mathbf{A}_N$ is a constant full rank weighting matrix $\mathbf{A}_N$. The GMM estimator of $\boldsymbol{\theta}$ is defined as $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} J_n(\boldsymbol{\theta})$ where $J_n(\boldsymbol{\theta}) \equiv (n^{-1}\sum_{i\in\mathcal{I}_n} \mathbf{z}_{n,i}\varepsilon_{n,i})^\top \mathbf{A}_n (n^{-1}\sum_{i\in\mathcal{I}_n} \mathbf{z}_{n,i}^\top \varepsilon_{n,i})$, the $(3K+1)\times(3K+1)$ full rank weighting matrix $\mathbf{A}_n$ is assumed to converge in-probability to $\mathbf{A}_N$. The linearity in (1.2) guarantees that the GMM estimator has a closed form solution given by

$$\widehat{\boldsymbol{\theta}}_{\text{GMM}} = [\mathbf{D}_n^\top \mathbf{Z}_n \mathbf{A}_n \mathbf{Z}_n^\top \mathbf{D}_n]^{-1}[\mathbf{D}_n^\top \mathbf{Z}_n \mathbf{A}_n \mathbf{Z}_n^\top \mathbf{y}_n]. \tag{1.4}$$

As to allow for the possibility that individuals' observed and unobserved characteristics to be correlated in the population joint distribution $\mathcal{F}(\mathbf{X}_N, \mathcal{G}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N)$, the concept of Doukhan and Louhichi's (1999) $\psi-$dependence in used here, see, e.g., Kojevnikov et al. (2020) and Estrada (2021). That is to bound the correlation between non-linear functions of random variables with the *dependence coefficients* which are decreasing functions of the network distance. From equation (1), the network $\mathcal{G}_{N,0}$ is independent of the regressors and the errors. Thus, we can rule out a dependence structure based on that network. However, because individuals endogenously form connections in $\mathcal{G}_N$ based on observed and unobserved characteristics, we would expect relatively high levels of dependence between individuals close to each other in the network space spanned by $\mathcal{G}_N$.

Following Estrada (2021), we define the random vector $\mathbf{r}_{N,i} \equiv [\mathbf{x}_{N,i}^\top, \varepsilon_{N,i}]^\top \in \mathbb{R}^{K+1}$. For $K, a \in \mathbb{N}_+$, endow $\mathbb{R}^{(K+1)\times a}$ 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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{(K+1)\times a}$. Let $\mathscr{L}_{K,a}$ denote the collection of bounded Lipschitz real functions mapping values from $\mathbb{R}^{(K+1)\times a}$ to $\mathbb{R}$. For each set of nodes A, let $\mathbf{r}_{N,A} = (\mathbf{r}_{N,i})_{i\in A}$, and we write triangular arrays simply as $\{\mathbf{r}_{N,i}\}$ and sequences say $\{\lambda_n\}_{n\geq 1}$ as $\{\lambda_n\}$ hereafter. Following Doukhan and Louhichi (1999) and Kojevnikov et al. (2020), $\psi-$dependence can be defined.

Definition 1 (ψ-dependence)**.** *A triangular array $\{\mathbf{r}_{n,i}\}$, $n \geq 1$, $\mathbf{r}_{n,i} \in \mathbb{R}^{K+1}$ is ψ-dependent if for each $n \in \mathbb{N}_+$ there exists a sequence $\{\lambda_n\} \equiv \{\lambda_{n,d}\}_{d\geq 0}, \lambda_{n,0} = 1$ 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 $g \in \mathscr{L}_{Q+1,b}$,*

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

The sequence $\{\lambda_n\}$ is called the *dependence coefficients* of $\mathbf{r}_{n,i}$. As mentioned before, the covariance of the non-linear functions of the random vectors $\mathbf{r}_{n,A}$ and $\mathbf{r}_{n,B}$ are bounded by the dependence coefficients $\lambda_{n,d}$ and a functional $\psi_{a,b}(f,g)$ that depends on the size of the sets A and B, and the aggregating non-linear functions f and g. Importantly by choosing appropriate values for the functions f and g, the ψ-dependence framework allows us to bound the dependence between observed and unobserved characteristics between any set of of individuals. As in Estrada (2021) we impose a week dependence assumption, but we do not impose a sharp bound on the decreasing pattern of the dependence coefficients with respect to the network distance.

Assumption 4 (Weak Dependence)**.** *Consider the set $\mathscr{G}$ of all possible realizations of $\mathcal{G}_N$ with positive probability mass in $\mathcal{F}$. For all networks $\mathcal{G}_N \in \mathscr{G}$, the conditional distribution $\mathcal{F}(\mathbf{X}_N, \boldsymbol{\varepsilon}_{\boldsymbol{N}} \mid \mathcal{G}_N)$ is such that:*

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

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

(iii) For each $n \in \mathbb{N}_+$, $\max_{d\geq 1} \lambda_{n,d} < \infty$ *and* $\lim_{d\to\infty} \lambda_{n,d} = 0$.

We impose Assumption 4(i) on the population, and it applies to the random vector $\mathbf{r}_{N,i} = [\mathbf{x}_{N,i}^\top, \varepsilon_{N,i}]^\top$. Given that any sample of size n is taken from the population characterized by the joint distribution $\mathcal{F}$, condition (i) applies for any $n \in \mathbb{N}_+$. Condition (ii) bounds the functional $\psi_{a,b}(f, g)$ by an arbitrary constant C, the cardinality of the sets A and B, the sup-norm and Lipschitz constants of the aggregating functions f and g. Intuitively, if the Lipschitz constants $\mathrm{Lip}(f)$ and $\mathrm{Lip}(g)$ increase, the values of the functions f and g can be larger for some values of $\mathbf{r}_{n,A}$ and $\mathbf{r}_{n,B}$, which requires larger constants to bound the covariance. The intuition is similar for the sup-norm. Finally, condition (iii) requires that the dependence coefficients are finite for any value of d, and that they dissipate to zero for large enough network distance between the random vectors $\mathbf{r}_{n,A}$ and $\mathbf{r}_{n,B}$.

The use of the $\psi-$dependence framework to modeling network dependence has the advantages that it does not impose functional form restrictions on the errors, and it allows correlation between indirectly connected nodes. However, transformations of ψ-dependent random variables are not necessarily ψ-dependent, therefore in order to analyze the asymptotic behavior of the estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$, we need to impose bounds to covariances of the form $\mathrm{cov}(r_{n,i,q}r_{n,j,\ell}, r_{n,h,q'}r_{n,s,\ell'})$, where $i, j, h, s \in \mathcal{I}_n$ and q, q', ℓ, and ℓ' are components of the vector $\mathbf{r}_{n,i}$. These include covariances such as $\mathrm{cov}(\varepsilon_{n,i}\varepsilon_{n,j}, \varepsilon_{n,h}\varepsilon_{n,s})$ or $\mathrm{cov}(x_{n,i,q}x_{n,j,\ell}, \varepsilon_{n,h}\varepsilon_{n,s})$ for example.

Assumption 5 (Bound Covariances)**.** *Define functions* $f_{q,\ell}$ *and* $g_{q',\ell'}$ *mapping* $\mathbb{R}^{(Q+1)\times 2}$ *into*

$\mathbb{R}$ *to 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* $q' \neq \ell'$. *The norms* $\|f_{q,\ell}(\mathbf{r}_{n,\{i,j\}})\|_{p_f^*} + \|g_{q',\ell'}(\mathbf{r}_{n,\{h,s\}})\|_{p_g^*} < \infty$ *for all* q, ℓ *where* $p_f^* = \max\{p_{f,i}, p_{f,j}\}$ *(analogous for* p_g^**) and* $1/p_{f,i} + 1/p_{f,j} + 1/p_{g,h} + 1/p_{g,s} < 1$.

The weak dependence Assumption in 4 guarantees that the dependence coefficients vanish to zero when the distance increases. However, the network distance $d_n(i, j)$ between any two individuals i and j is also a function of the sample size. Therefore, the asymptotic behavior of the dependence coefficients $\lambda_{n,d}$ depends on the asymptotic behavior of the network features determining the distance between nodes. In particular, the density of the network is explicitly linked with the geodesic distance. When the network density is arbitrarily large, the geodesic distance is always one for any pair of nodes. Therefore, as noted by Kojevnikov et al. (2020), there is a trade-off between network density and rate of convergence of the dependence coefficients. Networks with higher density would require the dependence to decrease faster (and vise-versa). The following assumption provides a necessary condition condition on the dependence coefficients for a Law of Large Numbers to apply.

Assumption 6 (Dependence Rate of Decay)**.** *Let* $\bar{D}_n(d) \equiv n^{-1}\sum_{i\in\mathcal{I}_n} |\mathcal{P}_n(i, d)|$ *be the average number of distance-d connections on the network* $\mathcal{G}_n$ *such that* $n^{-1}\sum_{d\geq 1} \bar{D}_n(d)\lambda_{n,d} \xrightarrow{a.s.} 0$ *as* $n \longrightarrow \infty$.

This Assumption is similar in spirit to Assumption 3.2 in Kojevnikov et al. (2020), but it includes the fact that in our setting the analyst has access to two types of networks that can involve weighted adjacency matrices. The following assumption imposes the existence of moments for products of ψ-dependent random variables, and it is a regularity condition needed for a Law of Large Numbers to apply.

Assumption 7 (Existence of Moments)**.** *$\exists\epsilon > 0$, $\sup_{n\geq 1}\max_{i\in\mathcal{I}_n}\|R_{n,i,j}\|_{1+\epsilon} < \infty$ a.s. where $R_{n,i,j} \equiv r_{n,i,q}r_{n,j,\ell}$ and $\|R_{n,i,j}\|_p \equiv (\mathbb{E}[|R_{n,i,j}|^p])^{1/p}$.*

The previous assumptions are enough to guarantee that the Law of Large Numbers applies for products of $\psi-$dependent random variables. To show asymptotic normality, we are using the Central Limit Theorem result in Kojevnikov et al. (2020). As mentioned before, for the asymptotic moments of network dependent random variables to be well defined, we need to control the level of asymptotic density. In particular, following Kojevnikov et al. (2020), we define a measure for the average neighborhood size as $\bar{D}_n(d,k) = 1/n\sum_{i\in\mathcal{I}_n}|\mathcal{P}_n(i,d)|^k$, and a measure for the average neighborhood shell size as

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

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

$$c_n(d,m,k) = \inf_{\alpha>1}\left[\bar{D}_n(d,m,k\alpha)^-\right]^{\frac{1}{\alpha}}\left[\bar{D}_n\left(d,\frac{\alpha}{\alpha-1}\right)\right]^{1-\frac{1}{\alpha}}. \tag{1.5}$$

For some arbitrary position q in the matrix $\mathbf{Z}_{n,i}$, let $S_n = \sum_{i\in\mathcal{I}_n} z_{n,i,q}\varepsilon_{n,i}$. Defining $\sigma_{n,q}^2 = \text{var}(S_n)$, the following assumption guarantees the existence of higher order moments, imposes asymptotic sparsity, and bound the long-run variance.

Assumption 8 (Average Sparsity)**.** *For all networks $\mathcal{G}_n \in \mathscr{G}$, (i) for some $p>4$, it follows that $\sup_{n\geq 1}\max_{i\in\mathcal{I}_n}\|z_{n,i,q}\varepsilon_{n,i}\|_p < \infty$. There exists a sequence $m_n \to \infty$ such that for $k =$*

$1, 2,$ *(ii)* , $\frac{n}{\sigma_{n,q}^{2+k}} \sum_{d \geq 0} c_n (d, m_n, k) \lambda_{n,d}^{1-\frac{2+k}{p}} \xrightarrow{a.s.} 0$ *as* $n \to \infty$, *(iii)* $\frac{n^2 \lambda_{n,m_n}^{1-(1/p)}}{\sigma_{n,q}} \xrightarrow{a.s.} 0$ *as* $n \to \infty$.

These conditions impose a rate of convergence of the dependence coefficients $\lambda_{n,d}$ that is related with the network density. There is a trade-off in which higher density requires higher speed in the dependence decreasing patterns. The previous assumptions are sufficient to show that our GMM estimator is consistent and asymptotically normal. Moreover, Lemma 3 in the 1.9 shows $\mathbf{\Omega}_n = \text{var}(\mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n)$ converges to the finite population variance

$$\mathbf{\Omega}_N = \lim_{n \to \infty} n^{-1} \left[\sum_{i=1}^{n} \text{var}(\mathbf{z}_{n,i} \varepsilon_{n,i}) + \sum_{i \neq j} \text{cov}(\mathbf{z}_{n,i} \varepsilon_{n,i}, \mathbf{z}_{n,j} \varepsilon_{n,j}) \right] \equiv \sum_{d \geq 0} \mathbf{\Gamma}_N(d) < \infty, \tag{1.6}$$

where $\mathbf{\Gamma}_N(d) = \sum_{i \in \mathcal{I}_N} \sum_{j \in \mathcal{P}_N(i,d)} \mathbb{E}[\mathbf{z}_{N,i} \varepsilon_{N,i} \varepsilon_{N,j} \mathbf{z}_{N,j}^\top]$ are the covariances for d. Therefore, the variance-covariance matrix $\mathbf{\Omega}_N$ can be calculated by summing the covariances for all possible distances $d \geq 0$. After characterizing $\mathbf{\Omega}_N$, Theorem 2 provides the asymptotic behaviour of (1.4).

Theorem 2. *Let Assumptions 2.1-8 hold, then as* $n \to \infty$, $\widehat{\boldsymbol{\theta}}_{\text{GMM}} = \boldsymbol{\theta} + o_p(1)$ *and* $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\text{GMM}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_N)$ *where the variance-covariance matrix is given by* $\mathbf{\Sigma}_N \equiv (\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top]^\top \mathbf{A}_N \mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top])^{-1} \times (\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top]^\top \mathbf{A}_N \mathbf{\Omega}_N \mathbf{A}_N \times \mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top])(\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top]^\top \mathbf{A}_N \mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top])^{-1}$, *and when* $\mathbf{A}_N = \mathbf{\Omega}_N^{-1}$ *then*

$$\mathbf{\Sigma}_N = (\mathbb{E}[\Sigma_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top]^\top \mathbf{\Omega}_N^{-1} \mathbb{E}[\Sigma_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top])^{-1}. \tag{1.7}$$

Recall that the expectation $\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i} \mathbf{d}_{N,i}^\top]$ can be written as $\boldsymbol{K}_{N,0,i} \mathbb{E}[\mathbf{z}_{N,i} \varepsilon_{N,i} \mid \boldsymbol{H}_{N,0,i}^* \neq \boldsymbol{O}_{pK}]$. Therefore, (1.7) depends on the population probabilities of an individual providing identification information. For low values of those probabilities, the upper right sub-matrix of $\boldsymbol{K}_{N,0,i}$ approaches the zero matrix, and the variance-covariance matrix could grow arbitrarily large. In the extreme case of non-identification, (1.7) diverges to infinity. Theorem 2 exposes a relationship between the network parameters' precision and network sparsity.

Efficient Weight Matrix Estimation

To construct the efficient version of the proposed GMM estimator we need a consistent estimator of $\boldsymbol{\Omega}_N$. Here we use Kojevnikov et al.'s (2020) network heteroscedasticity and autocorrelation-consistent (HAC) variance estimator. Let D_n represent a bandwidth after which the dependence between individuals vanishes. Then, the variance-covariance matrix estimator is given by

$$\widetilde{\boldsymbol{\Omega}}_n = \sum_{d \geq 0} K(d/D_n) \frac{1}{n} \sum_{i=1}^{n} \sum_{j \in \mathcal{P}_n(i,d)} \mathbf{z}_{n,i} \widetilde{\varepsilon}_{n,i} \widetilde{\varepsilon}_{n,j} \mathbf{z}_{n,j}^\top. \tag{1.8}$$

where $\widetilde{\varepsilon}_{n,i} = y_{n,i} - \mathbf{d}_{n,i}^\top \widetilde{\boldsymbol{\theta}}_{\text{GMM}}$, $K(\cdot)$ is the kernel function such that $K(0) = 1$ and $K(x) = 0$ for $x > 1$, and $\widetilde{\boldsymbol{\theta}}_{\text{GMM}}$ is a preliminary consistent estimator, e.g., (1.4) with $\mathbf{A}_n$ equal the identity matrix or $n^{-1}\mathbf{Z}_n^\top \mathbf{Z}_n$. In the second step, the feasible efficient GMM estimator is defined with $\mathbf{A}_n = \widetilde{\boldsymbol{\Omega}}_n^{-1}$ in (1.4), call it $\widehat{\boldsymbol{\theta}}^\star_{\text{GMM}}$.

Standard Error Calculation

It follows that the efficient variance-covariance matrix (1.7) can be estimated by

$$\left[n^{-1}\mathbf{D}_n^\top \mathbf{Z}_n \widehat{\boldsymbol{\Omega}}_n^{\star -1} n^{-1} \mathbf{Z}_n^\top \mathbf{D}_n\right]^{-1}, \tag{1.9}$$

where $\widehat{\boldsymbol{\Omega}}_n^{\star}$ is calculated as in (1.8) but using $\widehat{\boldsymbol{\theta}}_{\text{GMM}}$ instead. The standard errors can then be calculated by taking the squared-root of the main diagonal elements of (1.9) after dividing them by n.

1.5 Monte Carlo Experiments

As to showcase the versatility of the proposed estimator here, this section documents its performance in two different data generating processes (hereafter DGPs) where the endogeneity is generated by simultaneous determination of network formation and outcomes - also known as unobserved homophily- (Design 1) and measurement error in the connections (Design 2). A total of 1,500 data sets $\left\{y_i, x_i, \{w_{i,j}\}_{j=1,j\neq i}^n, \{w_{0;i,j}\}_{j=1,j\neq i}^n\right\}_{i=1}^n$ with $n \in \{50, 100, 200\}$ are generated from (2.1) by setting $k = 1$ and drawing $\{x_i\}_{i=1}^n$ as a random sample from a normal distribution with mean zero and variance 3. The other data components are constructed as follows:

Design 1: Unobserved Characteristics with Homophily

In this design, individual i's outcome variable, y_i, and connections $\{w_{i,j}\}_{j=1,j\neq i}^n$ are jointly determined through a common, $\varepsilon_{1;i}^*$, idiosyncratic homophily-related unobserved feature. Firstly, an exogenous adjacency matrix $\mathbf{W}_0 = [w_{0;i,j}]$ from a Erdös and Rényi's (1959) random graph with density 0.01 is generated along with a $n \times 1$ vector $\boldsymbol{\varepsilon}_1^* = [\varepsilon_{1;1}^*, \ldots, \varepsilon_{1;n}^*]^\top$ from a multivariate standard normal distribution. The elements of the endogenous adjacency matrix $\mathbf{W} = [w_{i,j}]$ are then calculated as

$$w_{ij} = \begin{cases} \mathbb{I}[|\varepsilon^*_{1;i} - \varepsilon^*_{1;j}| < \widehat{F}^{-1}_{\varepsilon^*_3}(0.95)] \times (1 - w_{0;i,j}) + w_{0;i,j} & ; \text{if } \varepsilon^*_{1;i} > \Phi^{-1}(0.95), \\ \mathbb{I}[|\varepsilon^*_{1;i} - \varepsilon^*_{1;j}| < \widehat{F}^{-1}_{\varepsilon^*_3}(0.95)] \times w_{0;i,j} & ; \text{if } \varepsilon^*_{1;i} < \Phi^{-1}(0.05), \\ w_{0;i,j} & ; \text{otherwise}, \end{cases}$$

where $\widehat{F}^{-1}_{\varepsilon^*_1}(0.95)$ represents the 95% empirical quantile of the elements of $\boldsymbol{\varepsilon}^*_1$, $\varepsilon^*_{1;k}$ represents its kth element, and $\Phi^{-1}(\cdot)$ represents the inverse of the cumulative distribution function of a standard normal random variable. This design captures the homophily idea, i.e., agents endowed with a large value of ε_1 will tend to create/mantain connections with those also endowed with large values of ε_1, and severe them with those with low values of this idiosyncratic unobserved feature. The $n \times 1$ vector of outcomes, $\boldsymbol{y}$, is then constructed from (2.1) by setting $\mathbf{v} = m \times \boldsymbol{\varepsilon_1} + \boldsymbol{\varepsilon_2}$, where $m \in \{1, 3\}$, $\boldsymbol{\varepsilon_2}$ is drawn from a multivariate standard normal distribution, and the elements of $\boldsymbol{\varepsilon}_1$ are defined as

$$\varepsilon_{1;i} = \begin{cases} \varepsilon^*_{1;i} & ; \text{if } \varepsilon^*_{1;i} < \Phi^{-1}(0.05) \text{ or } \varepsilon^*_{1;i} > \Phi^{-1}(0.95), \\ 0 & ; \text{otherwise}. \end{cases}$$

Design 2: Misclassified Links

This is a modified version of Lewbel et al.'s (2019) Monte Carlo design. While the true DGP involves an unobserved adjacency matrix $\mathbf{W}^*_0 = [w^*_{0;i,j}]$ generated from a standard Erdös and Rényi's (1959) random network model with density 0.01 of size n, the empiricist is assumed only to have access to an adjacency matrix, $\mathbf{W} = [w_{i,j}]$, with randomly misclassified links, i.e., $w_{i,j} = w^*_{0;i,j}e_{1;i,j} + (1 - w^*_{0;i,j})e_{2;i,j}$ for $i \neq j$, and an exogenous adjacency matrix $\mathbf{W}_0 = [w_{0;i,j}]$ where $w_{0;i,j} = w^*_{0;i,j}b_{1;i,j} + (1 - w^*_{0;i,j})b_{2;i,j}$ for $i \neq j$. The $e_{1;i,j}$, $e_{2;i,j}$, $b_{1;i,j}$, and $b_{2;i,j}$ are Bernoulli random variables drawn independently from each other $\forall i \neq j$ with parameters 0.5, 0, $1 - \tau$, and 0.002 respectively. The design parameter $\tau \in \{0.01, 0.05\}$ controls the probability of misclassification in $\mathbf{W}_0$. Notice that as in Lewbel et al. (2019), non-existing links are never misclassified in $\mathbf{W}$ but misclasification of these non-existing links are permitted in $\mathbf{W}_0$ with very small probability of 0.2%. However, this design makes the vector of individual outcomes an explicit function of the proportion of misclassification in $\mathbf{W}$ for each i, i.e., the $n \times 1$ vector $\mathbf{y}$ is constructed following equation (2.1) where $\mathbf{v} = \boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2$, $\varepsilon_{1,i} = 1/n \sum_{j=1}^{n} w^*_{0;i,j}e_{1;i,j}$, and $\boldsymbol{\varepsilon}_2$ is drawn from a multivariate standard normal distribution independently of everything else.

Results

Figures 1.1 and show the results in terms of box plots and Q-Q plots of the Monte Carlo replications. Apart from implementing the proposed GMM estimator in (1.4) which uses the 2SLS weighting matrix for $p = 2$ and $p = 3$, the performance of the standard Ordinary Least Squares (OLS) estimator, and the Generalized Two Stage Least Squares (G2SLS) estimator are also included. All adjacency matrices in all designs are row-normalized prior to estimation, see, e.g., Liu et al. (2014).

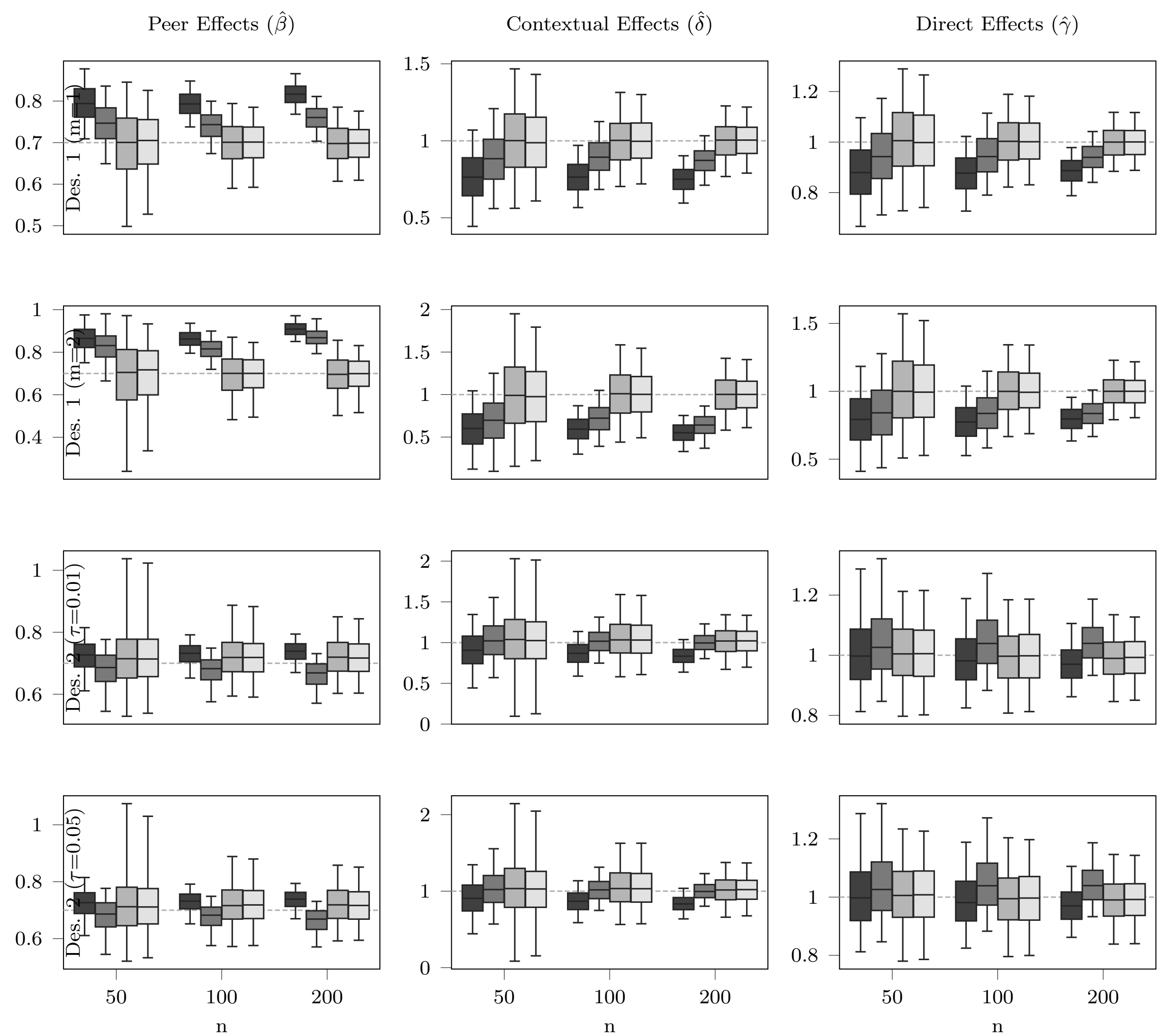
Each panel in Figure 1.1 displays the performance of the three estimators when the state of a design (Des.) changes by changing the relevant design parameter m or τ. The box plots based on the Monte Carlo replications of the OLS (black), G2SLS (dark gray), and the two proposed GMM estimators for $p = 2$ (gray) and $p = 3$ (light gray) of the social effects, i.e., peer effects (β), contextual effects (δ), and direct effects (γ) in (2.1) are shown with whiskers displaying the 5% and 95% empirical Monte Carlo quantiles. Across the board, for all parameters, designs, and sample sizes the proposed both GMM estimators performs better than the OLS and G2SLS estimators in terms of bias and sampling variability, and as expected, the estimation variability decreases when going from $p = 2$ to $p = 3$. In contrast, these results also show that the naive OLS and G2SLS could potentially lead to estimates with substantial biases in the presence of a endogenous network in a linear-in-means model. On average, the G2SLS underestimate the real value of the peer effects coefficient as shown in Chandrasekhar and Lewis (2016) for the case of missclassified links (Design 2).

Similarly, Figure displays the corresponding Q-Q plots for the G3SLS based on the standardized version of the Monte Carlo replications of the G3SLS estimator of the same social effects 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 4 works well even with a sample as small as 50 observations. Furthermore, as sample size increases the approximation improves for all parameters, and design parameters.

1.6 Application to Publication Outcomes in Economics

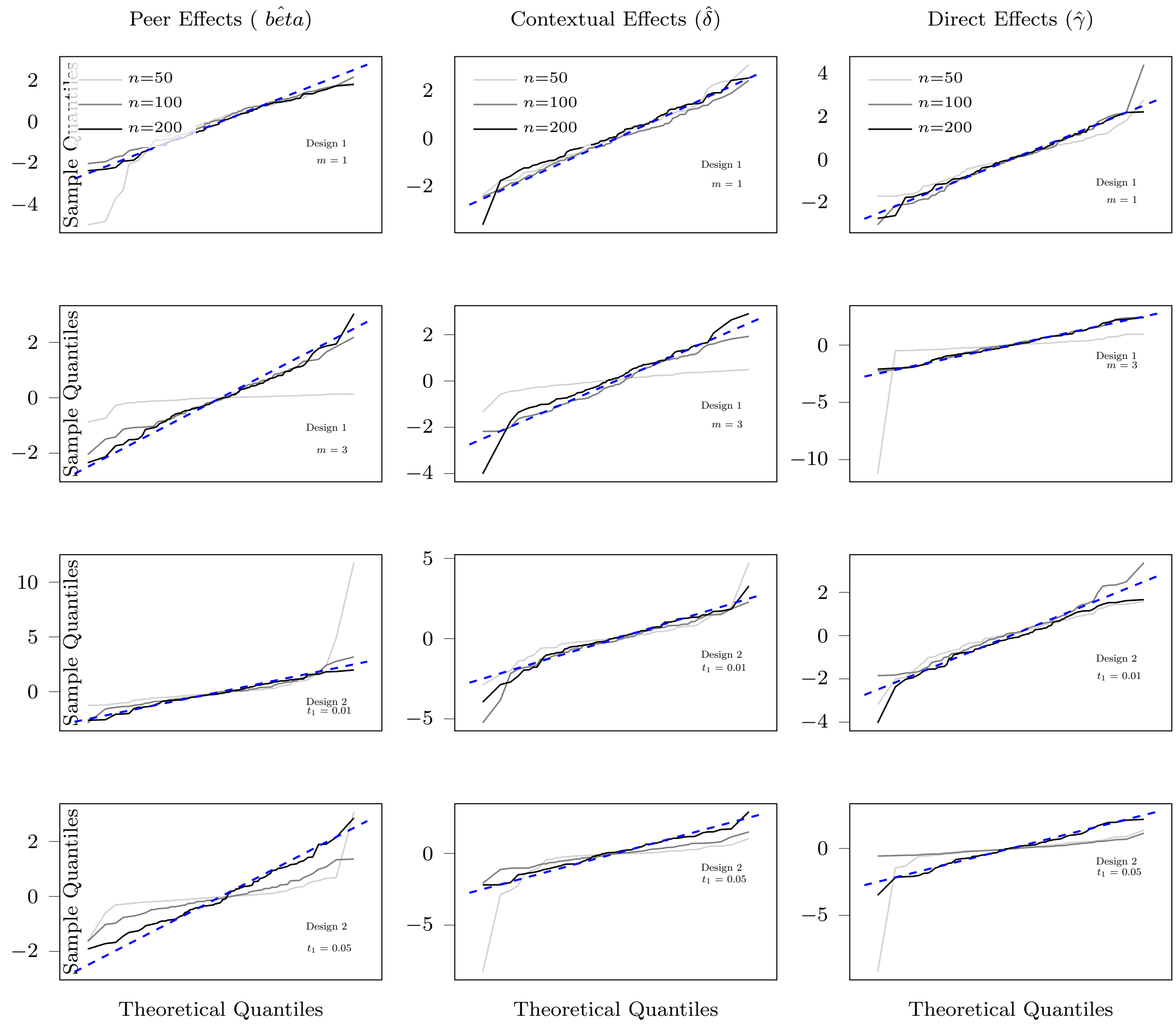
To illustrate the behavior of the efficient GMM estimator presented in Theorem 2, we use data of 1,628 peer-reviewed articles published between 2000-2006 in the top-four general-interest journals in Economics. The availability of online scientific research repositories has resulted

Figure 1.1: Box Plots of the OLS, G2SLS, and G3SLS Estimators of Social Effects



Note: Box plots depicts the Monte Carlo performance of the OLS (black), G2SLS (dark gray), and the proposed GMM estimators with the 2SLS weighting matrix for $p = 2$ (gray) and $p = 3$ (light gray) estimators of the parameters in (2.1) based on 1,500 replications of Design 1 (Des.1) and Design 2 (Des. 2) for sample sizes $n \in \{50, 100, 200\}$. The whiskers display the 5% and 95% empirical quantiles. The parameters m and τ control the level of endogeneity and the probability of misclassification in $\mathbf{W}_0$ respectively.

Figure 1.2: Q-Q Plots for the G3SLS Estimator of Social Effects



Note: Q-Q plots based on the standardized sample of 1,500 Monte Carlo replications of the proposed GMM estimator of the parameters in (2.1) for Design 1 (Des.1) and Design 2 (Des. 2) for sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black) and $p = 3$. The blue dashed line shows the 45 degree line. The parameters m and τ control the level of endogeneity and the probability of misclassification in $\mathbf{W}_0$ respectively.

in a stable source of data to uncover scholars' professional connections. Additionally, when linked with scholars' biographical public information, other type of professional connections beyond observed Co-authorship can be uncovered (Colussi, 2018). A major challenge when performing causal inference analysis is that co-authorship connections are inherently correlated with publications outcomes such as citation counts. However, the institutions where completed their graduate training although related to the co-authorship network, arguably does not a priori directly affect publications outcomes. This setting therefore directly lends itself to the use of the proposed estimator here to calculate social effects in research citations.

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' 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 here. 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 1.1 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.

Table 1.1: 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.

1.6.1 Multiplex Network Data

The multiplex network composed of two different types of connections among authors and editors (scholars hereafter) can be constructed using their co-authorship information, research interests, education, and employment history. In particular, the *Co-author* layer is made up of connections (edges) between scholars l and k if they co-authored a paper together. The *Alumni* layer is made of edges between scholars l and k if both obtained their Ph.D. from the same institution during the same time window. Details of how these two types of academic ties among scholars are constructed can be found in 1.6.1 below.

Since articles' authors are observed, these two types of professional ties can then be collapsed at the article level instead, i.e., articles i and j are connected in networks $\mathbf{W}$ or $\mathbf{W}_0$ if at least one of the authors of article i shares a co-authorship or alumni connection with at least one of the authors of article j respectively. Table 1.2 presents relevant network statistics for the implied articles' *Co-authors* ($\mathbf{W} = [\mathbf{w}_{i,j}]$) and *Alumni* ($\mathbf{W}_0 = [\mathbf{w}_{0;i,j}]$) networks. 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.

The two layers display low densities which is a common feature of empirical social networks (de Paula, 2017), with a somewhat small number of connected components for the *Alumni* network and a large number for the the *Co-authors* network. Although these networks are formed at the article level, and direct comparisons with classic collaboration network analysis as in Goyal et al. (2006) is not possible, they still display *small-world* properties, i.e., the levels of clustering (transitivity) are high while the average distance

Table 1.2: Network Statistics

Statistic	2000	2001	2002	2003	2004	2005	2006
Nodes	221	459	729	961	1,187	1,412	1,628
Co-authors Network							
Edges	28	189	674	1,217	2,120	3,302	4,732
Average Degree	0.25	0.82	1.85	2.53	3.57	4.68	5.81
Desity	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transitivity	0.82	0.83	0.71	0.66	0.65	0.63	0.6
Average Distance	1.04	1.10	1.25	1.57	1.92	2.56	3.32
Components	199	351	478	549	611	651	682
Alumni Network							
Edges	795	3,392	8,838	15,424	22,923	33,491	45,276
Average Degree	7.19	14.78	24.25	32.1	38.62	47.44	55.62
Desity	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Transitivity	0.57	0.57	0.55	0.55	0.55	0.54	0.53
Average Distance	3.14	3.25	2.95	2.80	2.81	2.73	2.70
Components	49	44	48	54	56	62	58
Same Faculty							

Note: 'Nodes' refers to the total number of articles including all previous years since 2000. 'Edges' refers to the total number of pair-wise connections among nodes. 'Average Degree' represents the average number of edges connected to each node, while 'Transitivity' presents the fraction of all possible triangles present in each network. 'Density' is defined as the ratio of the total number of observed edges to the total number of all possible edges in these networks. 'Components' displays the total number of connected components (subgraphs), while 'Average Distance' refers to the average number of steps along the shortest paths for all possible pairs of network nodes.

(average shortest path) is short (see, i.e., Humphries and Gurney, 2008, for a formal definition of a *small-world*).

Scholars Network Constructions

Co-authors – This type of connection happens when a scholar publishes a paper (alone or with other authors) in any of the four journals under consideration in any year of interest, and also publishes other papers (with the same or new 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.

Estimation of Network Effects for the Coauthors' Network

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 ($\ell = 1$) and advisorship ($\ell = 2$) are taken into account in the following specific case of the estimating equation (1.2):

$$y_{i,r,t} = \alpha + \beta \sum_{j \neq i} w_{i,j,t} y_{j,r,t} + \sum_{j \neq i} w_{i,j,t} \widetilde{\mathbf{x}}_{j,r,t}^{\top} \boldsymbol{\delta} + \mathbf{x}_{i,r,t}^{\top} \boldsymbol{\gamma} + \lambda_r + \lambda_t + \lambda_0 + v_{i,r,t}, \tag{1.10}$$

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 $w_{i,j,t}$ represents the (i, j) entry of the $\mathbf{W}$ adjacency matrix for the *co-authorship* network 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 network (`Isolated`). Contextual effects are only calculated for the `Editor` and `Different Gender` covariates, i.e., $\widetilde{\mathbf{x}}_{j,r,t}$. Estimating equation (1.10) also includes journal (λ_r), and year (λ_t). As mentioned before, the scalar structural error $v_{i,r,t}$ is such that $E_{\mathbf{X},\mathbf{W}}[\mathbf{v}] \neq 0$ because of the potential endogeneity of the co-authorship networks. 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 as in Designs 1 in Section 1.5.

Therefore, to be able to provide a causal interpretation for the peer effect parameters in equation (1.10), we need information on an additional network that is generated at random, or that can be considered good as random. In this chapter, we argue that the *Alumni* connections can be used as the exogenous network $\mathbf{W}_0$ in the identification Theorem 1. The main identifying assumption is that the natural timing when scholars form their *Alumni* connections make the network predetermined (and therefore uncorrelated with the unmeasurable characteristics of the contemporaneous outcome equation (1.10) such as professional recognition) because it was formed prior to the endogenous co-authorship connections, and by its nature to be correlated with them.

Model (1.10) 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 (1.10). Table 1.3 summarizes the results using the proposed G3SLS estimator while Tables 1.4 and 1.5 in 2.11 report the OLS and G2SLS results respectively. The GMM estimator uses the network HAC estimator of the covariance matrix in equation (2.5.1) where the function K is the Parzen kernel, and as suggested by Kojevnikov et al. (2020), we use the bandwidth $D_n = \text{ Cons. } \times [\log(\text{ avg.deg } \vee (1+\varepsilon))]^{-1} \times \log n$, where $Cons.$ equals 1.8, and $\epsilon = 0.05$, see Kojevnikov et al. (2020) for more details. For the both the OLS and G2SLS estimators the asymptotic standard errors are clustered at the corresponding network component level. This is a natural way of clustering when utilizing network data because each component corresponds to a portion of the network that are disconnected from each other, allowing for articles within each component to be correlated but not between disconnected components.

Table 1.3 provides different empirical results. Firstly, 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. This positive result can potentially reflect human capital

Table 1.3: Estimation Results for Social and Direct Effects

	Co-author Network				
	2002	2003	2004	2005	2006
Peer Effects ($\widehat{\beta}$)	0.604***	0.792***	0.906***	0.919***	0.732***
	(0.217)	(0.257)	(0.206)	(0.180)	(0.149)
Contextual Effects ($\widehat{\boldsymbol{\delta}}$)					
`Editor`	0.112	-0.501	-0.624*	-0.438	-0.433
	(0.419)	(0.355)	(0.377)	(0.378)	(0.268)
`Different Gender`	-1.921**	-1.156**	-0.852**	-0.567*	-0.224
	(0.775)	(0.497)	(0.392)	(0.302)	(0.361)
Direct Effects ($\widehat{\boldsymbol{\gamma}}$)					
`Editor`	-0.079	0.091	0.159	0.094	0.221*
	(0.21)	(0.173)	(0.162)	(0.148)	(0.114)
`Different Gender`	0.386***	0.348***	0.243***	0.233***	0.155*
	(0.132)	(0.111)	(0.098)	(0.082)	(0.085)
`Number of Pages`	0.027***	0.021***	0.018**	0.016***	0.018***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
`Number of Authors`	0.038	0.069	0.075**	0.086***	0.074***
	(0.051)	(0.042)	(0.037)	(0.032)	(0.027)
`Number of References`	0.008***	0.008***	0.008***	0.008***	0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
`Isolated`	1.608***	2.367***	2.763***	2.851***	2.260***
	(0.126)	(0.105)	(0.089)	(0.084)	(0.073)
n	729	961	1187	1412	1628
R^2	0.198	0.240	0.258	0.251	0.269

Note: Standard errors are in parenthesis and are calculated using the network HAC estimator of the covariance matrix in equation (2.5.1) where the function K is the Parzen kernel and the bandwidth D_n = Cons. $\times$ $[\log(\text{ avg.deg } \vee (1+\varepsilon))]^{-1} \times \log n$, where $Cons.$ equals 1.8 and $\epsilon = 0.05$. 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, Year and Alumni Network Components.

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. The `Editor` contextual effects from the coauthors' network are insignificant for all years. 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.

Similarly, a larger number of bibliographic references is associated with greater impact in terms of citations for the both network in all years for all estimators, while positive and significant effects in terms of number of authors are also found for all years for the advisorship network and in 2005 and 2006 for the co-authorship network for the proposed estimator. Finally, the direct effects of `Different Gender` on the articles' quality are all positive and significant for both networks and in all years. Numerically, the estimated values and implied asymptotic confidence intervals are similar across different networks. Across the board, 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 14% and 24% 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 (Card et al., 2020).

1.7 Conclusion

A novel, semiparametric, and computationally simple way to identify and consistently estimate social parameters in a linear-in-means model with endogenous network formation is proposed here. Identification can be achieved by the inclusion of the multiplex network data structure where at least one of the layers can be assumed to be exogenous (potentially predetermined), and the multiplex structure is such that the layers correlate with each other. This research shows that peer and contextual effects can be uniquely recovered from an estimating sample. Unlike current alternatives that require smoothing techniques and/or Bayesian methods, the resulting estimator is simple to compute utilizing any IV estimation routines in popular software like `Python`, `R`, or `Stata` for example. I establish the consistency and asymptotic normality of the proposed GMM estimator. I also characterize the form of the asymptotic variance-covariance matrix that accounts for the network dependence and illustrate how standard errors can be calculated in an empirical application.

An important aspect of the proposed methodology is that it recognizes that exogenously-imposed connections on individuals (such as randomization) do not necessarily cause social effects. However, they can generate new types of freely-formed connections that do so; i.e., resorting. The correlation between these two networks is at the heart of our identification and estimation strategy. In this sense, our approach provides an explicit solution to the resorting issue in random network identification strategies, such as in Moffitt (2001), by distinguishing what type of network creates peer effects (who you study with, for example) and what other type simply influences these connections, but are otherwise exogenous to the model (for instance, to whom your are randomly assigned to share a physical space). The type of endogeneity allowed in this framework is general enough to encompass settings with measurement error, sample selection, and correlation between the unobservables driving network formation and outcomes in a linear-in-means model. I present an empirical application where I use web scrapping and text mining tools to construct a data set consisting of all peer-reviewed research articles published in 4 of economics' top general-interest journals between 2000 and 2006. Using publicly available information of where the authors of these publications obtained their Ph.D. degree from, I construct the *Alumni* network and I argue that it is plausible to assume that the network is pre-determined yet correlated with the observed co-authorship ties among these scholars. 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.

Finally, another contribution of this research is technical in nature. A byproduct of acknowledging potential network mismeasurement or endogeneity is that it explicitly permits the observed and unobserved characteristics of individuals to be correlated; i.e., creating network dependence across observations in the sample. Our asymptotic results utilize the idea that dependence among observations decreases as a function of their distance in the network; i.e, $\psi-$dependence.

1.8 Appendix: Proofs of Main Results

Proof of Theorem 1. First note that Assumption 2 guarantees that the solution for model (1.2) exists. Assumption 1(*iii*) guarantees that the system of equations $\mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta})] = \mathbf{0}_K$ are not trivially satisfied by making $\eta_{N,0,i} = 0$ for all $i \in \mathcal{I}_N$. We show that the moment condition equation has a unique root at $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \boldsymbol{\delta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$. In particular, we show that there cannot be any other $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ different from $\boldsymbol{\theta}_0$ for which equation the moment condition holds. Choose an arbitrary vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ such that $\mathbb{E}[\mathbf{m}(\boldsymbol{\theta})] = 0$. Assumption 2 implies that $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}(y_{N,i} - \mathbf{d}_{N,i}^\top\boldsymbol{\theta})] = \mathbf{0}_K$. It follows that $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top](\boldsymbol{\theta}_0 - \boldsymbol{\theta}) + \mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\varepsilon_{N,i}] = \mathbf{0}_K$, and $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]\,(\boldsymbol{\theta}_0 - \boldsymbol{\theta}) = \mathbf{0}_K$. Under Assumption 3, it follows that $\mathbb{E}[\mathbf{m}(\boldsymbol{\theta})] = \mathbf{0}_K$ if and only if $\boldsymbol{\theta}_0 = \boldsymbol{\theta}$. □

Proof of Theorem 2. The GMM estimator in (1.4) in the main text can be written as

$$\widehat{\boldsymbol{\theta}}_{\text{GMM}} = \boldsymbol{\theta} + (n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\mathbf{A}_n n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n)^{-1}n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\mathbf{A}_n n^{-1}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n. \tag{1.11}$$

By construction, the matrix $\mathbf{A}_n$ is assume to converge to the full rank matrix $\mathbf{A}_N$ as $n \to \infty$. From Corollary 1, $n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n$ 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 3. Finally, Corollary 2 shows that $n^{-1}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n(\boldsymbol{\theta})$ converges to

$\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\theta})]=0$. It then follows that $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}=\boldsymbol{\theta}+o_p(1)$ as $n\to\infty$. For asymptotic normality, note that, from (1.11)

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

Let $\mathbf{Q}_{zx}=\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top}]$, then from Corollary 1 and Lemma 4, it follows that

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

and the result follows. The efficient variance-covariance matrix in (1.7) follows from standard matrix algebra calculations. □

1.9 Appendix: Proofs of Auxiliary Results

Lemma 1. *Let Assumptions 4 hold for $\{\mathbf{r}_{n,i}\}_{n\geq 1}$, $i\in\mathcal{I}_n$, and define $R_{i,j}=f_{q,\ell}(\mathbf{r}_{n,\{i,j\}})\equiv r_{n,i,q}r_{n,j,\ell}$ and $R_{h,s}=g_{q',\ell'}(\mathbf{r}_{n,\{h,s\}})\equiv r_{n,h,q'}r_{n,s,\ell'}$ 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 5 hold for $R_{i,j}$ and $R_{h,s}$, then*

$$|\mathrm{cov}\left(R_{i,j},R_{h,s}\right)|\leq 2\bar{\lambda}_{n,d}(C+16)\times 4\left(\pi_1+\tilde{\gamma}_1\right)\left(\pi_2+\tilde{\gamma}_2\right)\underline{\lambda}_{n,d}^{1-p_f-p_g}, \tag{1.12}$$

where $\underline{\lambda}_{n,d}=\lambda_{n,d}\wedge 1$, $\bar{\lambda}_{n,d}=\lambda_{n,d}\vee 1$, $\pi_1=\|\mathbf{r}_{n,i}\|_{p_{f,i}}\|\mathbf{r}_{n,j}\|_{p_{f,j}}$, $\pi_2=\|\mathbf{r}_{n,h}\|_{p_{f,h}}\|\mathbf{r}_{n,s}\|_{p_{f,s}}$, $\widetilde{\gamma}_1=\max\{\|\mathbf{r}_{n,i}\|_{p_{f,i}+p_{f,j}},\|\mathbf{r}_{n,j}\|_{p_{f,i}+p_{f,j}}\}$, $\widetilde{\gamma}_2=\max\{\|\mathbf{r}_{n,h}\|_{p_f},\|\mathbf{r}_{n,s}\|_{p_g}\}$ where $p_f=1/p_{f,i}+1/p_{f,j}$ and $p_g=1/p_{g,h}+1/p_{g,s}$, where the constant C is the same as in Assumption 4, the indexes i,j,h,s, and components q, q', ℓ, ℓ' may or may not be the same.

Proof. Define the increasing continuous functions $h_1(x)$ and $h_2(x)$ as in Theorem A.2 in

Kojevnikov et al. (Appendix A 2020, pp. 899-907) 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 5 guarantees the existence of the moments defining $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$. Then, Theorem A.2 in Kojevnikov et al. (Appendix A 2020, pp. 899-907) applies to this setting (see also Corollary A.2. in Appendix A in Kojevnikov et al., 2020, pp. 899-907). □

Lemma 2 (LLN for Products of ψ-dependent Random Variables)**.** *Let Assumptions 1 – 7 hold, define $R_{n,i,j} \equiv r_{n,i,q} r_{n,j,\ell}$ and form $\{R_{n,i,j}\}_{i \in \mathcal{I}_n, j \in \mathcal{I}_i}$, where $\mathcal{I}_i$ is a set of indexes defined for each $i \in \mathcal{I}_n$, then 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. Using the same approach of Jenish and Prucha (2009) and Kojevnikov et al. (2020), let the censoring function $\varphi_k(x) = (-K) \vee (K \wedge x)$ be such that, for some $k > 0$,

$$R_{n,i,j} = R^{(k)}_{n,i,j} + \tilde{R}^{(k)}_{n,i,j},$$

where $R^{(k)}_{n,i,j} = \varphi_k(R_{n,i,j})$, and $\tilde{R}^{(k)}_{n,i,j} = R_{n,i,j} - \varphi_k(R_{n,i,j}) = (R_{n,i,j} - \operatorname{sgn}(R_{n,i,j})k)\mathbb{1}\{|R_{n,i,j}| > k\}$. Let $\|X\|_k = (\mathbb{E}[|X|^k])^{1/k}$ for $k \in [1,\infty)$. Therefore, following the previous definition, and applying the triangle inequality

$$\begin{aligned}\left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w^*_{i,j} \left(R_{n,i,j} - \mathbb{E}\left[R_{n,i,j}\right]\right) \right\|_1 \leq& \left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w^*_{i,j} \left(R^{(k)}_{n,i,j} - \mathbb{E}\left[R^{(k)}_{n,i,j}\right]\right) \right\|_1 \\ &+ \left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w^*_{i,j} \left(\tilde{R}^{(k)}_{n,i,j} - \mathbb{E}\left[\tilde{R}^{(k)}_{n,i,j}\right]\right) \right\|_1.\end{aligned}$$

From Assumption 7, noting that the expectation on the second term of the previous equation is bounded by $\mathbb{E}[|\tilde{Y}^{(k)}_{n,i}|] = \mathbb{E}[|\tilde{Y}^{(k)}_{n,i}|\mathbb{1}\{|Y_{n,i}| > k\}] \leq 2\mathbb{E}[|Y_{n,i}|\mathbb{1}\{|Y_{n,i}| > k\}]$, and following the same arguments as in Kojevnikov et al. (2020) under which $\lim_{k\to\infty}\sup_{n\geq 1}\max_{i\in\mathcal{I}_n}\mathbb{E}\left[|R_{n,i,j}|\,1\left\{|Y_{n,i}|>k\right\}\right]=0$ a.s. By Lyapunov's inequality it follows that

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^*_{i,j}\left(R^{(k)}_{n,i,j}-\mathbb{E}\left[R^{(k)}_{n,i,j}\right]\right)\right\|_1 \leq \left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^*_{i,j}\left(R^{(k)}_{n,i,j}-\mathbb{E}\left[R^{(k)}_{n,i,j}\right]\right)\right\|_2, \tag{1.13}$$

where (1.13) is an expression for the standard deviation of $\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}R^{(k)}_{n,i,j}$. Note that

$$\text{var}\left(\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^*_{i,j}R^{(k)}_{n,i,j}\right) = \sum_{i\in\mathcal{I}_n}\text{var}\left(\sum_{j\in\mathcal{I}_i} w^*_{i,j}R^{(k)}_{n,i,j}\right) + \sum_{i\neq h\in\mathcal{I}_n}\text{cov}\left(\sum_{j\in\mathcal{I}_i} w^*_{i,j}R^{(k)}_{n,i,j}, \sum_{s\in\mathcal{I}_h} w^*_{h,s}R^{(k)}_{n,h,s}\right)$$

The variance part of the previous equation can be further expressed as

$$\begin{aligned}
\text{var}\left(\sum_{j\in\mathcal{I}_i} w^*_{i,j}R^{(k)}_{n,i,j}\right) &= \sum_{j\in\mathcal{I}_i} w^{*2}_{i,j}\text{var}(R^{(k)}_{n,i,j}) + \sum_{j\neq s\in\mathcal{I}_i} w^*_{i,j}w^*_{i,s}\text{cov}(R^{(k)}_{n,i,j}, R^{(k)}_{n,i,s}) \qquad (1.14)\\
&\leq C\sum_{j\in\mathcal{I}_i} w^{*2}_{i,j} + \sum_{j\in\mathcal{I}_i}\sum_{d\geq 1}\sum_{s\in\mathcal{P}_n(j,d)\cap\mathcal{I}_i} |\text{cov}(R^{(k)}_{i,j}R^{(k)}_{i,s})|\\
&\leq C\sum_{j\in\mathcal{I}_i} w^{*2}_{i,j} + \psi_{1,1}\left(\varphi_k,\varphi_k\right)\sum_{d\geq 1}\lambda_{n,d}\sum_{j\in\mathcal{I}_i}\left|\mathcal{P}_n(j,d)\right|,
\end{aligned}$$

where the second inequality follows form $w^*_{i,j}, w^*_{i,s} \in [0,1]$. In the first term of the second inequality, C represents any generic constant from the fact that after the initial partition of $R_{n,i,j}$, the variance of $R^{(k)}_{n,i,j}$ is bounded. The last inequality follows from two reasons.

Firstly, $|\text{cov}(R^{(k)}_{i,j}R^{(k)}_{i,s})| \leq \psi_{1,1}(\varphi_k, \varphi_k)\,\lambda_{n,d}$ for $d_n(i,j) = d$ and φ_k a bounded function with $\text{Lip}\,(\psi_k) = 1$. Secondly, the fact that the set of indexes $\mathcal{P}_n(j,d) \cap \mathcal{I}_i \subset \mathcal{P}_n(j,d)$. The covariance component can be written as

$$\begin{aligned} \text{cov}\left(\sum_{j\in\mathcal{I}_i} w^*_{i,j}R^{(k)}_{n,i,j}, \sum_{s\in\mathcal{I}_h} w^*_{h,s}R^{(k)}_{n,h,s}\right) &= \sum_{j\in\mathcal{I}_i}\sum_{s\in\mathcal{I}_h} w^*_{i,j}w^*_{h,s}\text{cov}(R^{(k)}_{n,i,j}, R^{(k)}_{n,h,s}) \\ &\leq \sum_{j\in\mathcal{I}_i}\sum_{d\geq 1}\sum_{s\in\mathcal{P}_n(j,d)\cap\mathcal{I}_h} |\text{cov}(R^{(k)}_{i,j}R^{(k)}_{h,s})| \\ &\leq \psi_{1,1}(\varphi_k, \varphi_k)\sum_{d\geq 1}\lambda_{n,d}\sum_{j\in\mathcal{I}_i}|\mathcal{P}_n(j,d)|\,, \end{aligned} \tag{1.15}$$

where the second and third inequalities follow from the same principles discussed before. It follows from equations (1.14) and (1.15) that the total variance of $\sum_{j\in\mathcal{I}_i} w^*_{i,j}R^{(k)}_{n,i,j}$ can be bounded by

$$\begin{aligned} \text{var}\left(\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^*_{i,j}R^{(k)}_{n,i,j}\right) &= C\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^{*2}_{i,j} + 2\psi_{1,1}(\varphi_k, \varphi_k)\sum_{i\in\mathcal{I}_n}\sum_{d\geq 1}\lambda_{n,d}\sum_{j\in\mathcal{I}_i}|\mathcal{P}_n(j,d)| \\ &= C\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^{*2}_{i,j} + 2\psi_{1,1}(\varphi_k, \varphi_k)\sum_{d\geq 1}\lambda_{n,d}\sum_{i\in\mathcal{I}_n}|\mathcal{P}_n(j,d)| \\ &\leq n\left(C\bar{\mathcal{I}}_n + 2\psi_{1,1}\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d}\right), \end{aligned} \tag{1.16}$$

where $\bar{\mathcal{I}}_n = n^{-1}\sum_{i\in\mathcal{I}_n}|\mathcal{I}_i|$ and the inequality follows because $w^{*2}_{i,j} \in [0,1]$. The set $\mathcal{I}_i$ can either be empty, equal to the union of individual i's degree in the networks $\mathcal{G}$ and $\mathcal{G}_0$, or equal to $\mathcal{P}_n(i,1)$ (individual i's degree in network $\mathcal{G}$). Note that $|\mathcal{I}_i| \leq |\mathcal{P}_n(i,1)|$ for all i. Also,

$\sum_{i \in \mathcal{I}_n} |\mathcal{P}_n(i,1)|\lambda_{n,1} \leq \sum_{d \geq 1} \bar{D}_n(d)\lambda_{n,d}$, which converges to zero almost surely by Assumption 6. It follows that $n^{-1}\bar{\mathcal{I}}_n \xrightarrow{\text{a.s.}} 0$. Therefore,

$$\left\| \frac{1}{n} \sum_{i \in \mathcal{I}_{n,m}} \sum_{j \in \mathcal{I}_i} w^*_{i,j} \left(R^{(k)}_{n,i,j} - \mathbb{E}\left[R^{(k)}_{n,i,j} \right] \right) \right\|_1 \leq \left(n^{-1} C \bar{\mathcal{I}}_n + 2\psi_{1,1} n^{-1} \sum_{d \geq 1} \bar{D}_n(d)\lambda_{n,d} \right)^{1/2}. \quad (1.17)$$

The result follows from $n^{-1}\bar{\mathcal{I}}_n \xrightarrow{\text{a.s.}} 0$ and $n^{-1}\sum_{d \geq 1} \bar{D}_n(d)\lambda_{n,d} \xrightarrow{\text{a.s.}} 0$ under Assumption 6.

□

Corollary 1 (LLN for Instruments and Regressors). *Let Assumptions 1 to 6 hold. Then,*

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

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 $\mathbf{w}_{n,0,i}\mathbf{x}_{n,q}\mathbf{w}_{n,i}\mathbf{x}_{n,\ell}$, (3) network and non-network regressors of the form $\mathbf{w}_{n,0,i}\mathbf{x}_{n,q}\mathbf{x}_{n,i,\ell}$, and (4) network regressors and network outcomes of the form $\mathbf{w}_{n,0,i}\mathbf{x}_{n,q}\mathbf{w}_{n,i}\mathbf{y}_n$ (and the versions of (2) and (3) with $\mathbf{w}^p_{n,0,i}$ instead of $\mathbf{w}_{n,0,i}$). The LLN follows from Lemma 2 by choosing $\mathcal{I}_i =$ for (1), $\mathcal{I}_i$ to be the union of individual i's degree in the networks $\mathcal{G}$ and $\mathcal{G}_0$ in (2), and $\mathcal{I}_i = \mathcal{P}_n(i,1)$ for (3). For (4) note that

$$\mathbb{E}[\mathbf{W}_N \mathbf{y}] = \gamma_0 \mathbf{W}_N \mathbf{x}_N + (\gamma_0 \beta_0 + \delta_0) \sum_{p=0}^{\infty} \beta_0^k \mathbf{W}_N^{p+2} \mathbf{x}_N, \quad (1.18)$$

and by choosing $\mathcal{I}_i$ to be the union of individual i's degree in the network $\mathcal{G}$ and the set of individuals at distance p from i (for all $p \in \mathbb{R}_+$), Lemma 2 applies for all the values in the infinite sum formed by $\mathbf{w}_{n,0,i}\mathbf{x}_{n,q}\mathbf{w}_{n,i}\mathbf{y}_n$ after replacing $\mathbf{w}_{n,i}\mathbf{y}_n$ from equation (1.18) (the same argument holds for (2) and (3) when using $\mathbf{w}^p_{n,0,i}$ instead of $\mathbf{w}_{n,0,i}$) . 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_0 from Assumption 2, completing the proof. □

Corollary 2 (LLN for Instruments and Errors)**.** *Let Assumptions 1 to 6 hold, then*

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

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

Lemma 3 (Finite Variance)**.** *Define* $\mathbf{S}_n = \mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n$ *and* $\boldsymbol{\Omega}_n = \text{var}(n^{-1/2}\mathbf{S}_n)$ *and let Assumptions 1 to 6 hold, then as* $n \to \infty$, $\boldsymbol{\Omega}_n \xrightarrow{a.s.} \boldsymbol{\Omega}_N < \infty$.

Proof. As before $n^{-1/2}\mathbf{S}_n \equiv n^{-1/2}\sum_{i=1}^n \mathbf{z}_{n,i}\varepsilon_{n,i}$. The bounded covariance assumptions from Lemma 1 combined with the arguments in Lemma 2 guarantee that $\lim_{n\to\infty} n^{-1}\text{var}\left(\sum_{i=1}^n \mathbf{z}_{n,i}\varepsilon_{n,i}\right)$ is finite. In particular, from equation (1.17), using the appropriate values for $R_{n,i,j}$ and $\mathcal{I}_i$, $n_{m,\lambda}$ (see Corollary 1), it follows that $\text{var}(\sum_{i=1}^n \mathbf{z}_{n,i}\varepsilon_{n,i}) = O_p(1)$. Given that $\boldsymbol{\Omega}_n$ converges to a finite quantity, it follows that $\boldsymbol{\Omega}_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}_N$, where

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

□

Lemma 4 (Central Limit Theorem)**.** *Let Assumptions 1 to 8 hold and define* $S_n \equiv \sum_{i \in \mathcal{I}_n} z_{n,i,q}\varepsilon_{n,i}$*, where* $z_{n,i,q}$ *is the* qth *entrance of the vector* $\mathbf{z}_{n,i}$*. Then by definition of* $\mathbf{z}_i$*,* $\mathbb{E}[z_{n,i,q}\varepsilon_{n,i}] = 0$*, and 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 $\sigma_n \equiv var(S_n)$ *and* $\Phi(\cdot)$ *denotes the cumulative distribution function of a standard normal random variable.*

Proof. Let $Y_{n,i} = z_{n,i,q}\varepsilon_{n,i}$, from Lemma 1, 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 Kojevnikov et al. (2020, Appendix A, pp. 899-907) to $Y_{n,i}$ and S_n/σ_n, respectively. □

Lemma 5 (Multivariate Central Limit Theorem)**.** *Under Assumptions 1 to 8 hold, then as* $n \to \infty$*,* $n^{-1/2}\sum_{i=1}^{n} \mathbf{z}_{n,i}\varepsilon_i \xrightarrow{d} \mathcal{N}(0, \Omega_N)$*.*

Proof. From Lemma 4 it follows that $n^{-1/2}\sum_{i=1}^{n} z_{n,i,q}\varepsilon_{n,i} \xrightarrow{d} \mathcal{N}(0, \sigma_n^2)$, while from Lemma 3 it follows that $\mathbf{\Omega}_N$ exists. Therefore, the result follows from an application of the Cramér-Wold device. □

1.10 Appendix: Robustness and Additional Empirical Results

Table 1.4: Estimation Results for Social and Direct Effects using the OLS Estimator

	Co-author Network				
	2002	2003	2004	2005	2006
Peer Effects ($\widehat{\beta}$)	0.375***	0.414***	0.474***	0.470***	0.471***
	(0.064)	(0.052)	(0.047)	(0.045)	(0.039)
Contextual Effects ($\widehat{\boldsymbol{\delta}}$)					
`Editor`	0.266	0.133	0.004	-0.056	-0.020
	(0.259)	(0.163)	(0.148)	(0.144)	(0.130)
`Different Gender`	-0.087	0.036	-0.040	0.008	-0.013
	(0.190)	(0.130)	(0.116)	(0.110)	(0.103)
Direct Effects ($\widehat{\gamma}$)					
`Editor`	0.039	-0.021	-0.024	0.054	0.083
	(0.136)	(0.117)	(0.127)	(0.122)	(0.117)
`Different Gender`	0.216*	0.183*	0.179**	0.133*	0.123
	(0.118)	(0.103)	(0.088)	(0.079)	(0.078)
`Number of Pages`	0.028***	0.025***	0.021***	0.018***	0.017***
	(0.053)	(0.044)	(0.039)	(0.034)	(0.029)
`Number of Authors`	0.081	0.094**	0.085**	0.092***	0.073**
	(0.057)	(0.047)	(0.041)	(0.036)	(0.030)
`Number of References`	0.010***	0.010***	0.009***	0.010***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
`Isolated`	1.105***	1.243***	1.355***	1.322***	1.320***
	(0.124)	(0.104)	(0.096)	(0.085)	(0.082)
n	729	961	1187	1412	1628
R^2	0.259	0.292	0.298	0.283	0.282

Note: Estimation results using the OLS estimator. Standard errors are in parenthesis and are clustered at the specific network's components. 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, Year and Alumni Network Components.

Table 1.5: Estimation Results for Social and Direct Effects using the G2SLS Estimator

	Co-author Network				
	2002	2003	2004	2005	2006
Peer Effects ($\widehat{\beta}$)	0.665***	0.604***	0.456***	0.505***	0.600***
	(0.163)	(0.113)	(0.079)	(0.085)	(0.112)
Contextual Effects ($\widehat{\boldsymbol{\delta}}$)					
`Editor`	0.061	0.020	0.001	-0.077	-0.056
	(0.295)	(0.189)	(0.153)	(0.150)	(0.137)
`Different Gender`	-0.198	-0.032	-0.067	-0.028	-0.075
	(0.217)	(0.150)	(0.131)	(0.123)	(0.121)
Direct Effects ($\widehat{\boldsymbol{\gamma}}$)					
`Editor`	-0.032	-0.041	-0.032	0.038	0.070
	(0.156)	(0.126)	(0.129)	(0.129)	(0.123)
`Different Gender`	0.244*	0.192*	0.190**	0.143*	0.134*
	(0.131)	(0.111)	(0.091)	(0.081)	(0.081)
`Number of Pages`	0.027***	0.024***	0.021***	0.018***	0.016***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)
`Number of Authors`	0.082	0.092**	0.077*	0.085**	0.065**
	(0.056)	(0.046)	(0.041)	(0.036)	(0.030)
`Number of References`	0.008***	0.009***	0.009***	0.010***	0.010***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
`Isolated`	2.139***	1.936***	1.299***	1.447***	1.788***
	(0.131)	(0.109)	(0.099)	(0.089)	(0.087)
n	729	961	1187	1412	1628
R^2	0.286	0.320	0.337	0.319	0.311

Note: Standard errors are in parenthesis and are clustered at the specific network's components. 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, Year and Alumni Network Components.

Chapter 2

Estimation of Multilayered Networks Effects with Observational Data

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. I use a multilayer network data structure to characterize my data generating process and accommodate multiple social networks. My methodology allows all layers in the multilayer network to be endogenous, which is fundamental when dealing with observational data. I show that identification of heterogeneous network effects is possible under the assumption that the dependence between individuals in the population is characterized by a ψ*-dependent stochastic process, which guarantees that their dependence vanishes in the network space. I offer a novel multilayer measure of distance that, combined with the* ψ*-dependence assumption, provides a source of exogenous variation that I use to form identifying moment conditions. I propose a Generalized Method of Moments estimator that is consistent and asymptotically normal at the standard rate. I characterize* *the asymptotic variance-covariance matrix that considers the intrinsic network dependence among individuals. I show that too dense or too sparse networks provide weak identifying information that translates into larger standard errors. A Monte Carlo experiment confirms the desirable finite properties of the proposed estimator. An empirical application finds positive and significant peer effects in citations from a multilayer network of professional connections among scholars publishing in top general interest journals in economics.*

2.1 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 requires an explicit network formation model, see, e.g., Johnsson and Moon (2019) and references therein. 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).

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.

As in Bramoullé et al. (2009), 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, I assume that a $\psi-$dependent stochastic process characterizes the dependence of individuals in the population (Doukhan and Louhichi, 1999; Kojevnikov et al., 2020). 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 me to construct moment conditions to separately identify the potential M different peer and contextual effects after controlling for the presence of correlated effects (endogenous network formation).

I propose an innovation to the idea of ψ-dependence to accommodate multilayer network data structures by proposing 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 me to take 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, I use 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 could 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 article 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, Estrada et al. (2020) 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. (2021) provides identification results for a MLiM model with exogenous network formation. The results in this paper are more general because I do not require any layer in the multilayer network data to be strictly exogenous, and the proposed estimator collapses to the estimators in Estrada et al. (2020) and Manta et al. (2021) when additional exogeneity assumptions are imposed in the model.

I model the potential network endogeneity 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 introduce the *weak dependence* 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. I use these conditions to show that the parameters of the MLiM model are point identified.

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. I study limiting distributions when $n \to \infty$. The linearity assumption of the MLiM model guarantees that the resulting GMM estimator has a closed-form solution. My 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 in the population. I use a HAC estimator of the variance-covariance matrix 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. I use the multilayer network data to uncover positive and significant peer effects in citations from the co-authorship network among articles published by these scholars. However, I do not find peer effects from any of the other types of networks included in the estimating model. 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. 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. I also find positive results of research teams that are gender diverse on the quality of a paper measured in terms of citation outcomes after controlling for other articles' characteristics such as number of pages, number of bibliographic references, and various network fixed effects.

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 Estrada et al. (2020). 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 importantly separates from the structural endogeneity approach. It does not have to assume a particular source of unobserved heterogeneity, and it does not require network formation estimation. The framework also separates from the literature using natural or artificial experiments to randomize peers because I explicitly assume that the networks can form endogenously. Thus, my approach is closer to the literature using random shocks (experimental or quasi-experimental) on the regressors for identification. The reason 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 2.2 introduces the multilayer network data structure and its representation in terms of adjacency matrices. Section 2.3 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 2.1 presents conditions for the parameters of the MLiM to be uniquely recovered (point identification) from the joint distribution characterizing the infinite population of interest, and outlines a network formation model under which the main identifying assumptions are satisfied. Section 2.5 describes the proposed GMM estimation procedure, its asymptotic distribution, and how to calculate valid asymptotic standard errors. Section 2.6 presents a Monte Carlo simulation study, while section 2.7 presents the empirical application to publication outcomes in economics. Finally, Section 2.8 concludes. The supplemental materials contain all mathematical proofs of the main results, the proofs for intermediate results, proposed algorithms, and data construction in 2.9, 2.10 and Appendix 2.11, respectively.

2.2 Background

This section introduces the background and notation necessary to develop the framework for identification and estimation. Following Boccaletti et al.'s (2014) notation, a multi-layer network is a pair $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ where $\mathcal{G} = \{G_m; m \in \{1, \ldots, M\}\}$ 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 $\mathcal{C} = \{E_{m,s} \subseteq V_m \times V_s; m, s \in \{1, \ldots, M\}, m \neq s\}$. 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.

2.3 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 could be derived as a best response function of an structural game of social interactions with a quadratic utility (Blume et al., 2015). However, I take the MLiM model as my primitive 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. I argue that it is possible to write the optimal choice (outcome) of an individual as

$$y_{N,i} = \alpha^0 + \sum_{m=1}^{M} \sum_{j \neq i} w_{N,m;i,j} y_{N,j} \beta_m^0 + \sum_{m=1}^{M} \sum_{j \neq i} w_{N,m;i,j} \mathbf{x}_{N,j}^\top \boldsymbol{\delta}_m^0 + \mathbf{x}_{N,i}^\top \boldsymbol{\gamma}^0 + \varepsilon_{N,i}, \qquad (2.1)$$

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). 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 $[\beta_1^0, \dots \beta_M^0, \boldsymbol{\delta}_1^0, \dots, \boldsymbol{\delta}_M^0, \alpha^0, \boldsymbol{\gamma}^0]$ is the true parameter vector. The model can be written in matrix form as

$$\mathbf{y}_N = \alpha^0 \boldsymbol{\iota}_N + \left(\sum_{m=1}^{M} \beta_m^0 \mathbf{W}_{N,m} \right) \mathbf{y}_N + \sum_{m=1}^{M} \mathbf{W}_{N,m} \mathbf{X}_N \boldsymbol{\delta}_m^0 + \mathbf{X}_N \boldsymbol{\gamma}^0 + \boldsymbol{\varepsilon}_N. \tag{2.2}$$

This article focuses on the representation of the multilayer network as the set of layers' adjacency matrices. The reason is that the proposed MLiM model does not consider interlayer connections to affect individuals' outcomes. The multilayer measure of distance define in section 2.1 can handle data structures where the multilayer network could contain interlayer connections. Then, so long as the model in (2.2) 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 (2.2) 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 6 in 2.10 shows that the parametric restrictions in Assumption 9 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 6 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 9 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 9 (Invertibility)**.** *The adjacency matrix of layer m has no cycles, i.e.,* $w_{N,ii} = 0$ *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* $\|\mathbf{W}_{N,m}\|_\infty = \sup_i \sum_{j=1}^N |w_{N,m;ij}|$ *and* $m = 1, \ldots M$*.*

If Assumption 9 is satisfied, it follows that the solution for equation (2.2) can be written in terms of $\mathbf{X}_N$, $\boldsymbol{\varepsilon}_N$, and $\mathcal{M}_N$ as follows

$$\mathbf{y}_N = \mathbf{S}^{-1}(\boldsymbol{\beta}^0, \mathcal{M}_N)\left(\alpha^0 \boldsymbol{\iota}_N + \sum_{m=1}^{M} \mathbf{W}_{N,m}\mathbf{X}_N\boldsymbol{\delta}_m^0 + \mathbf{X}_N\boldsymbol{\gamma}^0 + \boldsymbol{\varepsilon}_N\right), \tag{2.3}$$

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, i.e., $\mathbb{E}(\boldsymbol{\varepsilon}_N|\mathbf{X}_N, \mathbf{W}_{N,1}, \ldots, \mathbf{W}_{N,M}) \neq 0$. 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

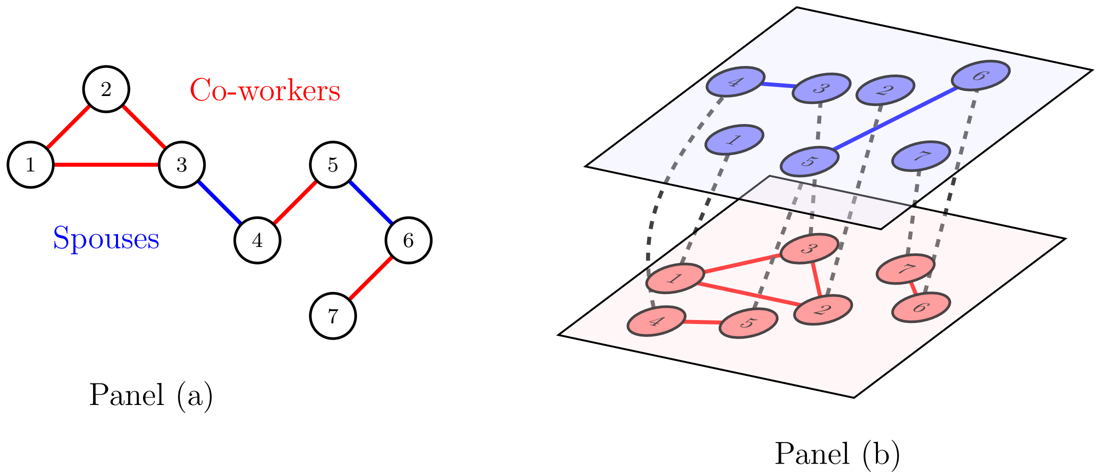
$\boldsymbol{\varepsilon}_N$. Recent work investigating the estimation of social effects under network endogeneity has been based on approaches that explicitly model the network formation process, see, e.g., Johnsson and Moon (2019). This article, however, abstract away from explicit network formation assumptions. Instead, as I detail in section 2.1, the main idea to overcome this endogeneity issue is to form independent multilayer sub-network structures that allow me to create moment conditions to identify the parameters in equation (2.3) 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.

2.3.1 Examples

Example 1 (non-overlapping networks): De Giorgi et al. (2020) studies consumption network effects in a context where co-workers are the relevant reference group. Acknowledging the potential network endogeneity, the authors use 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 2.1 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

Figure 2.1: De Giorgi et al.'s (2020) Data Structure Represented as a Multilayer Network

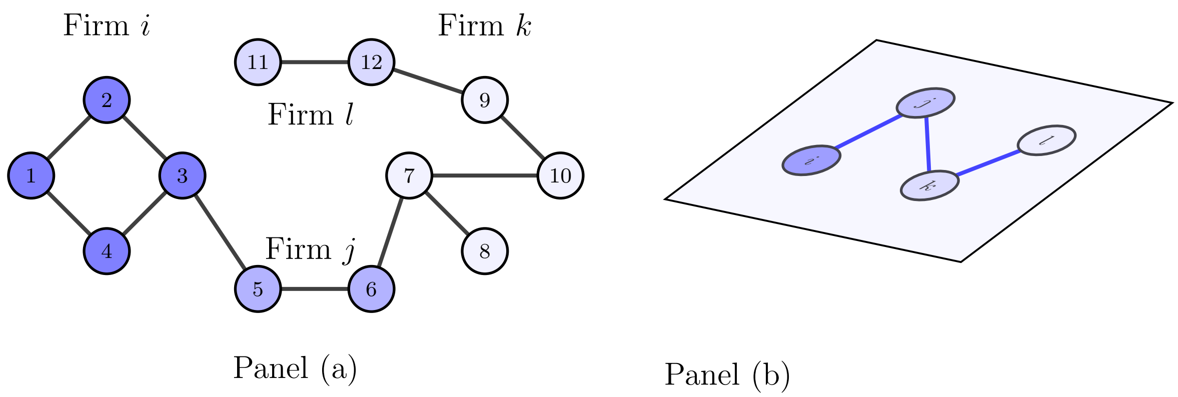


Note: Panel (a) displays an example of the non-overlapping network data structure in De Giorgi et al. (2020) which is a modification of the example presented in Figure 2 of their article. The blue edges represent co-worker's connections while the red edges illustrate spouses relationships. Panel (b) 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.

node in one layer to itself in the other layer. 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.

Example 2 (monolayer network): Zacchia (2019) studies 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 2.2 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

Figure 2.2: 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.

as a multilayer network by setting $M = 1$.

2.4 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 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.

2.4.1 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 $d^*_N(i,j)$. I use the star superscript 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, my 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, define the set of all possible shortest paths 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 2.1 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

$$d_N^{\mathcal{M}}(i,j) = d_N^*(i,j) + \tau_{i,j} c_N^*(i,j), \tag{2.4}$$

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

$$d_N(A,B) = \min_{i \in A} \min_{j \in B} d_N^{\mathcal{M}}(i,j). \tag{2.5}$$

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 I will use in the following sections.

Definition 2 (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 $\mathcal{P}_n^+(i,d,m)$.*

2.4.2 Network Dependence

This section introduces a helpful framework to characterize the levels of network dependence between the regressors and the errors in equation (2.1). 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 $\mathbf{r}_{A,N} = (\mathbf{r}_{N,i})_{i\in A}$. I accommodate the ψ-dependence definitions in Doukhan and Louhichi (1999) and Kojevnikov et al. (2020) to my multilayer network framework.

Definition 3 (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 exists 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 2.5. 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 3 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 10 (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 $C > 0$*,* $\psi_{a,b}(f,g) \leq C \times ab\,(\|f\|_\infty + \text{Lip}(f))\,(\|g\|_\infty + \text{Lip}(g))$*.*

(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 $\mathcal{F}(\mathbf{X}_N, \boldsymbol{\varepsilon_N} \mid \mathcal{M}_N)$. The idea is that any endogenous process of network formation will not place correlated individuals close to each other. Subsection 2.4.3 presents a network formation process under which the WND assumption holds. Part (ii) of Assumption 10 states that the functional bounding the covariance in definition 3 is increasing in in the set sizes, the sup-norm of the aggregating functions f and g, and their Lipschitz constants, $\text{Lip}(f)$ and $\text{Lip}(g)$, 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 me to form 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

[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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{(Q+1)\times a}$.

the regressors.

Assumption 11 (Errors' Moments)**.** *The unobserved shocks* $\varepsilon_{N,i}$ *are such that (i)* $\mathbb{E}(\varepsilon_{N,i}) = 0$ *and (ii)* $\mathbb{E}\left(\mathbf{x}_{N,i}\varepsilon_{N,i}\right) = \mathbf{0}_{Q\times 1}$ *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 11 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), Estrada et al. (2020) 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 11 in the sense that they think of the network formation process as generating the dependence between individuals' observed and unobserved attributes, see e.g, Johnsson and Moon (2019) and Auerbach (2022). The following proposition characterizes the identifying moment conditions.

Proposition 1. *Let Assumptions 10 and 11 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*

$$\mathbb{E}\left(\mathbf{x}_{N,j}\varepsilon_{N,i} \mid c_N^*(i,j) \geq K_c, d_N^c(i,j) \geq K_d\right) = \mathbf{0}_{Q\times 1}, \tag{2.6}$$

$$\mathbb{E}\left(\mathbf{x}_{N,j}\varepsilon_{N,i} \mid c_N^*(i,j) < K_c, d_N^*(i,j) \geq K_d\right) = \mathbf{0}_{Q\times 1}. \tag{2.7}$$

2.9 presents the proof for Proposition 1. The proof shows that if $K_d \leq K_c + 1$, the conditioning set in equation (2.6) does not provide different information from that in equation (2.7). 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 the values of K_c and K_d completely characterizes the measure of distance $d_N^{\mathcal{M}}(i,j)$. The object $d_N^c(i,j)$ captures the second shortest path between individuals i and j for which the number of changes in edge types is lower than $c_N^*(i,j)$. If no path exists with those characteristics, then $d_N^c(i,j) \to \infty$. The object $d_N^c(i,j)$ is vital to guarantee that a second longer path with lower edge type changes does not connect individuals also connected through the shortest path with enough edge type changes. For instance, in Figure 2.1, $d_{1,5}(2) \to \infty$, because there is not path connecting 1 and 5 that has less than two edge type changes. 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 (2.6) 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

(2.7) 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 2.1 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 (2.6) 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 2.1 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 (2.6) holds. This property makes

[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.

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 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 (2.7) in this paper (see Section 2.1 in Zacchia, 2019, pp. 1994). Given a monolayer data structure su particular information structure (see Assumption 2 and Proposition 2 in Section 2.1 in Zacchia, 2019, pp. 1995-1997).

2.4.3 Example of a Network Formation Model

As mentioned before, the estimator identification and estimation approaches I propose in this paper, do not require explicit assumptions on the network formation process. However, a specific network formation rule under which Assumption 10 holds can be useful to check the plausibility of higher level conditions. Thus, this section outlines a network formation model under which the WND Assumption, and consequently, the moment conditions in Proposition 1, are justified. The central insight comes from the small world model proposed by Jackson and Rogers (2005), see also Jackson (2010), where an exogenous group of clusters (also known as islands) determines the individuals' incentives to form connections. Individuals are more likely to create links within than across islands. From a statistical perspective, in this example, the clusters impose restrictions on the joint distribution of the regressors and the errors. One sufficient condition to imply WND is to assume the existence of a collection of random variables determining link formation that can be arbitrarily dependent intracluster but not intercluster. This assumption can be thought as imposing a sharp bound on the dependence between individuals who are sufficiently spread out as measured by their characteristics (Kuersteiner, 2019; Kuersteiner and Prucha, 2020).

Individuals are exogenously distributed across different clusters. I think of this process as an endowment to individuals with initial characteristics that feature local dependence (for instance, being born in a wealthy family or geographical endowments). Individuals then form connections in the multilayer network in the context of a link announcement game where the payoffs are functions of the random variables generated across different clusters. Let q_i and ν_i be observed and unobserved (to the econometrician) random variables affecting the position of individual i in the multilayer network. These random variables are drawn from an arbitrary joint distribution for all individuals in the same cluster but independently across different clusters. Individuals' utility functions include homophily taste on q_i and ν_i, and it may include payoff externalities (Miyauchi, 2016; Mele, 2017; de Paula et al., 2018; Christakis et al., 2020; Sheng, 2020), or degree heterogeneity (Graham, 2017). Given that my goal in this section is not to identify the parameters of the network formation model but to characterize the statistical dependence generated by the network structure, I can be agnostic about the specifics of the utility function. In particular, for the sake of simplicity, I assume that only homophily effects are relevant to characterize individuals linking behavior.

A critical difference between the existing network formation models and my proposed setup is the multilayer network data structure. As in Joshi et al. (2020a), I argue that links across different networks exhibit complementarities. Individuals experience higher marginal utilities when they connect otherwise disconnected components with a link of a different type. 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). Following Joshi et al. (2020b), I assume an exhaustive game of multilayer formation, which is a sequential game that is played one layer at a time via a link announcement game. The sequence of games repeat until agents have no incentives to delete or form new links, i.e., pairwise stable multilayer network (Jackson and Wolinsky, 1996). One simple network formation rule that encompasses the strategic features described before is given by

$$w_{ij,m} = \mathbb{1}\{\pi_g|g_i - g_g| + \pi_\nu|\nu_i - \nu_j| + \pi_C \mathbb{1}\{C_{i,-m} \neq C_{j,-m}\} + u_{ij,m} > 0\}, \tag{2.8}$$

where the first two components, with $\pi_g, \pi_\nu < 0$, capture the homophily effect: individuals are more likely to form a connection if they are more similar in observed and unobserved characteristics, the third component captures the increasing marginal utility of forming bridge links with individuals who belong to disconnected components that do not include edges from the m layer, and $u_{ij,m}$ is a dyadic, layer specific shock.

The distances between the pairs of random vectors (g_i, g_j) and (ν_i, ν_j) when i and j belong to the same cluster are more likely to be short when the within-cluster dependence features positive correlations. In that case, individuals within the same cluster are more likely to be connected. Moreover, component $\mathbb{1}\{C_{i,-m} \neq C_{j,-m}\}$ in the marginal utility makes it more likely that edge-type changes happen across clusters. Under this network formation model, it is possible to determine the minimum number of edges required to guarantee traveling from one cluster to another by using paths with a given number of edge type changes. Because the random vectors (g_i, ν_j) and (g_j, ν_j) are independent for any i and j in different clusters, it follows that Assumption 10 holds for this simple network formation model.

Moreover, with the knowledge of the cluster membership, it is also possible to calculate the values of K_c and K_d that satisfy equations (2.6) and (2.7). Endogeneity will occur if the random variables g_i and ν_i are correlated with the regressors and the errors in the outcome equation (2.1). Given my emphasis on non-experimental data, endogeneity is the more likely scenario in this context. Of course, this is not the only network formation model that can rationalize the high-level conditions in Assumption 10. I include this discussion to justify the WND conditions and present a plausible set of individual incentives under which those assumptions are fulfilled. I also emphasize that the network formation model is required for neither identification nor estimation, as the primary goal of this paper is to propose a robust method to arbitrary network endogeneity.

2.4.4 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 (2.6) or (2.7) 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 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 12 (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 (2.1) 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 $\mathbf{D}_N = \left[\mathbf{W}_{N,1}\mathbf{y}_N, \ldots, \mathbf{W}_{N,M}\mathbf{y}_N, \mathbf{W}_{N,1}\mathbf{X}_N, \ldots, \mathbf{W}_{N,M}\mathbf{X}_N, \widetilde{\mathbf{X}}_N\right]$ associated with the vector $\boldsymbol{\psi} = [\beta_1, \ldots, \beta_M, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_M, \widetilde{\gamma}]$, 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 $\mathbf{Z}_N = \left[\mathcal{W}_{N,1,\beta}\mathbf{X}_N, \ldots, \mathcal{W}_{N,M,\beta}\mathbf{X}_N, \mathcal{W}_{N,1,\delta}\mathbf{X}_N, \ldots, \mathcal{W}_{N,M,\delta}\mathbf{X}_N, \widetilde{\mathbf{X}}_N\right]$. 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 $\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi})] = \mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi}) \mid \mathcal{H}^*_{N,\lambda,i} \neq \mathbf{O}_{R\times R}] + (\mathcal{I}_{1+R+Q} - \mathcal{K}_{N,\lambda,i})\mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}(\boldsymbol{\psi}) \mid \mathcal{H}^*_{N,\lambda,i} = \mathbf{O}_{R\times R}]$, 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 $\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top}] = \mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top} \mid \mathcal{H}_{N,\lambda,i}^{*}, \mathcal{H}_{N,i}^{*} \neq \mathbf{O}_{R\times R}]\mathcal{K}_{N,i}$, where $\mathcal{K}_{N,i}$ 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 $\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top}]$ cannot have full column rank. The following assumption imposes a column rank condition on the matrix $\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top}]$.

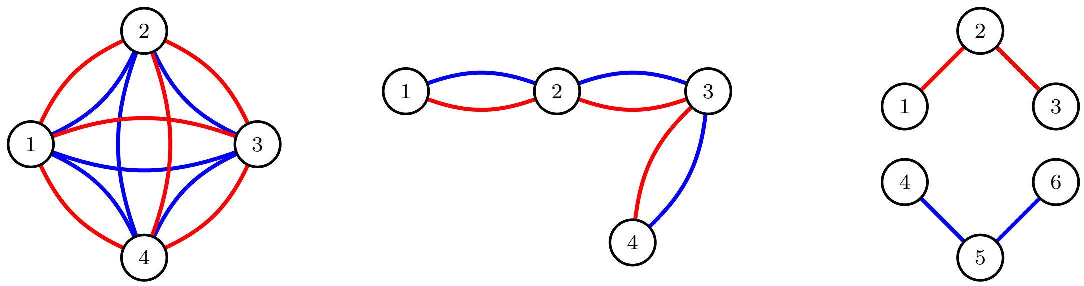
Assumption 13 (Relevance)**.** *The matrix* $\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top}] < \infty$ *has full column rank.*

As mentioned above, parts (i) and (ii) of Assumption 12 are necessary conditions for Assumption 13 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 $\mathbf{Z}_n^{\top}\mathbf{D}_n$ 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.

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 13 holds for given values of K_c with K_d. Figure 2.3 presents three examples where Assumption 13 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 13 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 13 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 13 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 (2.2). The conditions in Assumption 9 required for invertibility also

Figure 2.3: Failure of Assumption 13



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.

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 $m \neq s$ as shown in the Remark 2 in 2.10. In the monolayer network case, it is well known 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). 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 2.10 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 3. *Suppose Assumptions 9, 10, 11, 12, and 13 hold for some K_c and K_d such that $K_d > K_c + 1$. Then, the parameters $\boldsymbol{\psi}^0 = [\beta_1^0, \ldots, \beta_M^0, \boldsymbol{\delta}_1^0, \ldots, \boldsymbol{\delta}_M^0, \alpha^0, \boldsymbol{\gamma}^0]$ 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 3 is based on the existence of a set of regressors for which Assumption 11 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 11 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).

2.9 presents the proof for Theorem 3. 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 3, I now 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 (2.1) reduces to

$$y_{N,i} = \alpha + \sum_{j \neq i} w_{N,1;i,j} y_j \beta + \sum_{j \neq i} w_{1;i,j} \mathbf{x}_{N,j}^{\top} \boldsymbol{\delta} + \mathbf{x}_{N,i}^{\top} \boldsymbol{\gamma} + \varepsilon_{N,i}, \tag{2.9}$$

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 (2.6) 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 2.1:

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

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 $\mathbb{E}\left[\mathcal{W}_{N,1,\beta}\mathbf{x}_N\boldsymbol{\varepsilon}_N\right]$ produces a vector of expectations with components such as $\mathbb{E}\left[x_4\varepsilon_1\right]$, $\mathbb{E}\left[x_5\varepsilon_1\right]$, $\mathbb{E}\left[x_6\varepsilon_1\right]$, $\mathbb{E}\left[x_7\varepsilon_1\right]$, 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 $\mathbf{Z}_N = \left[\mathbf{W}_{N,1}\mathbf{W}_{N,2}\mathbf{X}_N, \mathbf{W}_{N,1}\mathbf{X}_N, \widetilde{\mathbf{X}}_N\right]$. Theorem 3 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 2.7. 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 2.2, this matrix is given by

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

Based on $\mathcal{W}_{N,\delta}$, the matrix of instruments can be constructed as $\mathbf{Z}_N = \left[\mathcal{W}_{N,\delta}\mathbf{X}_N, \widetilde{\mathbf{X}}_N\right]$, and $\mathbf{D}_N = \left[\mathbf{W}_N\mathbf{X}_N, \widetilde{\mathbf{X}}_N\right]$. Theorem 3 can be also apply for this case to show that the parameters $[\boldsymbol{\delta}, \alpha, \boldsymbol{\gamma}]$ are point identified.

2.5 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 3 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 (2.1). 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 annalist 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 (2.1). 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 (2.6) and (2.7) 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 Appendix 2.11 and may be of independent interest for those interested in path problems.

I form the GMM estimator from the moment condition $J_N(\boldsymbol{\psi}) = \mathbb{E}[\mathbf{m}_N(\boldsymbol{\psi})]^\top(\mathbf{A}_N^\top\mathbf{A}_N)\mathbb{E}[\mathbf{m}_N(\boldsymbol{\psi})]$, 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 (2.1) 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 2.4.1, the week dependence Assumption in 10 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 14 (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 $f_{q,\ell}(\mathbf{r}_{n,\{i,j\}}) = r^q_{n,i} r^\ell_{n,j}$ and $g_{q',\ell'}(\mathbf{r}_{n,\{h,s\}}) = r^{q'}_{n,h} r^{\ell'}_{n,s}$ 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 2, let $\mathcal{P}_n(i, 1, m, \lambda) = \{i : i, j \in \mathcal{I}_N$, and $w_{n,m\lambda;i,j} > 0\}$.

Assumption 15 (Bounded Weights)**.** *For all networks $\mathcal{M}_n \in \mathscr{M}$, (i) for all lay-*

ers 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 15 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 2.3, 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 16 (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 16 captures the trade off between the network density and the level of dependence that the model can allows via the dependence coefficients $\theta_{n,d}$. Intuitively, it guarantees that, on average, the level of sparsity capture by $\bar{D}_n(d)$ does not increase faster than the dependence between individuals at distance d for all possible distances. Assumption 16 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 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

$$c_n(d,m,k) = \inf_{\alpha>1}\left[\delta_n^-(d,m,k\alpha)\right]^{\frac{1}{\alpha}}\left[\delta_n\left(d;\frac{\alpha}{\alpha-1}\right)\right]^{1-\frac{1}{\alpha}}. \tag{2.10}$$

The measure of average density in equation 2.10 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 bound the long-run variance.

Assumption 17 (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 10 and 11 in 2.10 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 13 shows that $\boldsymbol{\Omega}_n = \mathrm{Var}(\mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n)$ converges to the population variance

$$\boldsymbol{\Omega}_N = \lim_{n\to\infty} n^{-1}\left[\sum_{i=1}^{n} \mathrm{var}(\mathbf{z}_{n,i}\varepsilon_{n,i}) + \sum_{i\neq j} \mathrm{cov}(\mathbf{z}_{n,i}\varepsilon_{n,i}, \mathbf{z}_{n,j}\varepsilon_{n,j})\right] < \infty, \tag{2.11}$$

which is necessary for the multivariate central limit theorem result in Lemma 14. 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 (2.11) can be written in terms of a generic distance d as

$$\boldsymbol{\Gamma}_N(d) = \sum_{i\in\mathcal{I}_N} \sum_{j\in\mathcal{P}_N(i,d)} \mathbb{E}\left[\mathbf{z}_{N,i}\varepsilon_{N,i}\varepsilon_{N,j}\mathbf{z}_{N,j}^\top\right], \tag{2.12}$$

which implies that the population variance-covaraince matrix $\boldsymbol{\Omega}_N$ can be constructed as the sum of the covariance estimators in equation (2.12) for all possible distances $d \geq 0$ (equality is allowed to account for the variance),

$$\mathbf{\Omega}_N = \sum_{d \geq 0} \mathbf{\Gamma}_N(d). \tag{2.13}$$

After characterizing the variance-covariance matrix $\mathbf{\Omega}_N$ both in terms of individual and distances sums, Theorem 4 shows that the optimal GMM estimator is consistent and asymptotically normal. The optimal estimator is given by $\widehat{\boldsymbol{\psi}}^{\star}_{GMM} = (\mathbf{D}_n^{\top}\mathbf{Z}_n\mathbf{\Omega}_N^{-1}\mathbf{Z}_n^{\top}\mathbf{D}_n)^{-1}(\mathbf{D}_n^{\top}\mathbf{Z}_n\mathbf{\Omega}_N^{-1}\mathbf{Z}_n^{\top}\mathbf{y}_n)$.

Theorem 4. *Let Assumptions 9-16 hold, then* $\widehat{\boldsymbol{\psi}}^{\star}_{GMM} = \boldsymbol{\psi} + o_p(1)$ *and* $\sqrt{n}(\widehat{\boldsymbol{\psi}}^{\star}_{GMM} - \boldsymbol{\psi}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^{\star}_N)$*, where* $\mathbf{\Sigma}^{\star}_N = (\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top}]^{\top}\mathbf{\Omega}_N^{-1}\mathbb{E}[\sum_{i \in \mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^{\top}])^{-1}$.

2.9 presents the proof for Theorem 4. 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. I interpret the results in this theorem 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, $\boldsymbol{\Sigma}_N^\star$ 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 $\widehat{\boldsymbol{\psi}}_{GMM}^\star$ 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 2.3 for an example). Theorem 4 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.

2.5.1 Covariance Matrix Estimation

The estimation of the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}_N^\star$ 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 10). 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

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

where $\widehat{\varepsilon}_{n,i} = y_{n,i} - \mathbf{d}_{n,i}^\top \widehat{\boldsymbol{\psi}}_{GMM}^\star$, 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

$$\widehat{\mathbf{\Sigma}}_n^\star = \left[n^{-1}\mathbf{Z}_n^\top \mathbf{D}_n \widehat{\mathbf{\Omega}}_n^{-1} n^{-1}\mathbf{Z}_n^\top \mathbf{D}_n \right]^{-1}. \tag{2.14}$$

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 $\widehat{\mathbf{\Sigma}}_n^\star$ after dividing them by n. The estimation of the efficient weighting matrix in (2.14) requires a consistent estimator of $\boldsymbol{\psi}$. I use a standard two-step GMM approach 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.

2.6 Monte Carlo Simulations

This section presents the simulation experiments designed to test the finite sample properties of the GMM estimator proposed in Theorem 4 and its robustness to endogenous multilayer network formation. To that end, I use a data generating process that mirrors the endogenous multilayer formation rule described in section 2.4.3. In particular, I define the number of regressors in equation (2.1) as $Q = 1$ and assume that the same observed characteristic x_i affects both the outcome in (2.1) and the network formation rule in (2.8). I set the number of exogenous clusters to be $K = 10$, and randomly separate all nodes into these clusters. For each cluster $c \in \{1, \ldots, 10\}$, I draw the unobserved random vector $\mathbf{g}_c$ of size n_c from a multivariate normal with mean μ_c, variance 1, and intercluster correlation of 0.9 for all clusters c. 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, I generate the vector of observed characteristics for cluster c as $\mathbf{x}_c = \mathbf{g}_c + \boldsymbol{\epsilon}_{1,c}$, where $\boldsymbol{\epsilon}_{1,c}$ 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 (2.8). I approximate the exhaustive iterative process by performing six iterations of the link announcement game. Based on the resulting network, I construct the outcome vector $\mathbf{y}$ following equation (2.3) 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 I know the cluster membership within the data generating process (DGP), and I created the 10 clusters independent to each other, I can then exactly calculate the values of K_c and K_d that make equations (2.6) and (2.7) 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, I simulated the DGP until I reached a total of 1,000 data sets.

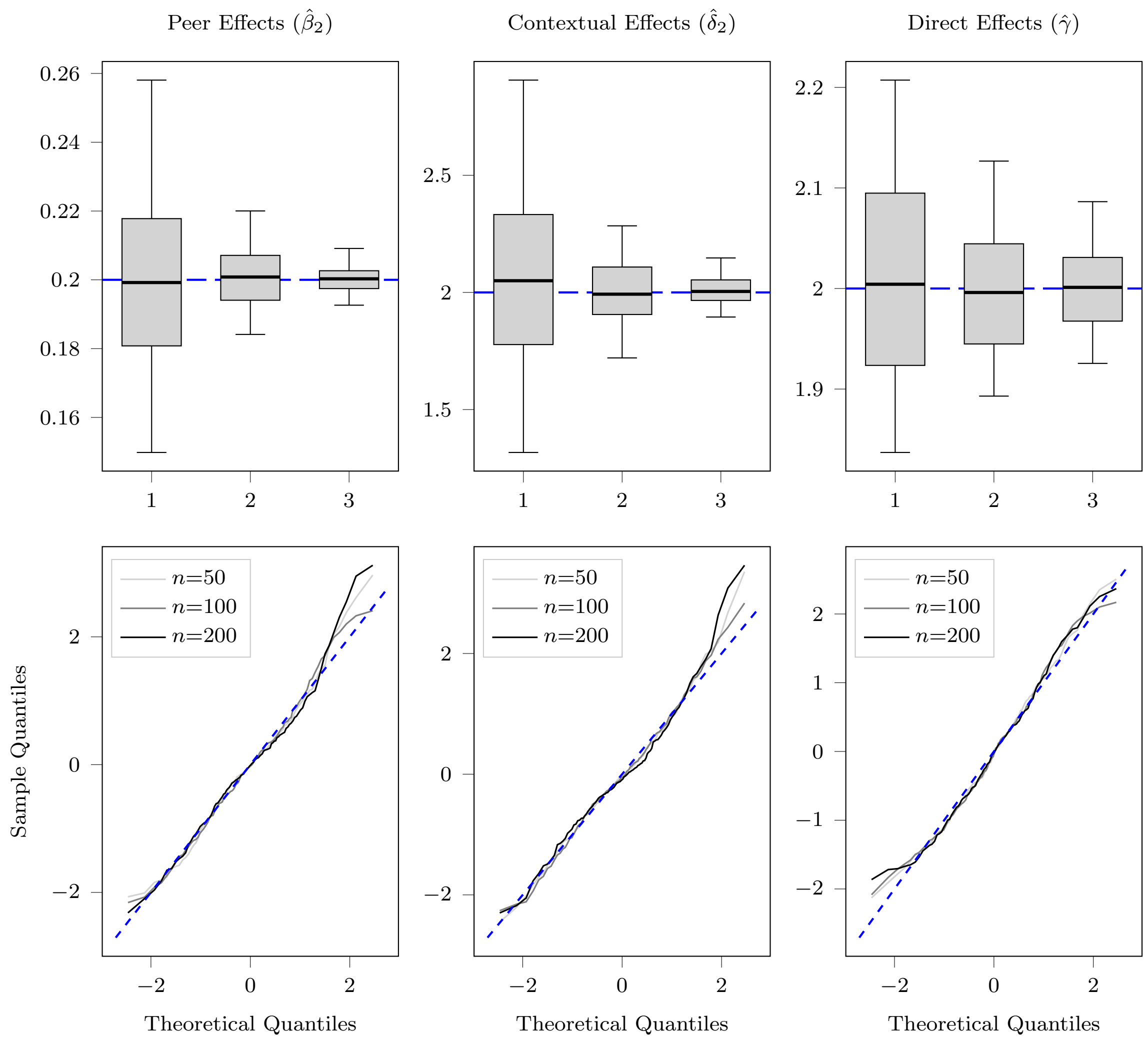
Based on the 1,000 data sets $\left\{y_{n;i}, x_{n;i}, \{w_{n,1;i,j}\}_{j\neq i}^{n}, \{w_{n,2;i,j}\}_{j\neq i}^{n}, \{w_{n,3;i,j}\}_{j\neq i}^{n}\right\}_{i=1}^{n}$ with $n \in \{50, 100, 200\}$, I estimate the parameters of interest $[\beta_1^0, \beta_2^0, \beta_3^0, \delta_1^0, \delta_2^0, \delta_3^0, \gamma^0, \alpha^0]$ by using the proposed GMM estimator. Given that $Q = 1$, in this case, the system is just identified, and I can disregard concerns about efficiency. Figure 2.4 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 Appendix 2.12. The first row in Figure 2.4 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 2.4 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 4 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.

2.7 Application to Publication Outcomes in Economics

As in Chapter 1 Section 2.7, 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. Differing from the previous Chapter, this empirical application includes four different types of networks: co-authorship ($m = 1$), alumni ($m = 2$), advisorship ($m = 3$), and colleagues ($m = 4$). I specify how to construct the co-authorship and alumni networks in the previous Chapter. The other two networks are constructed as follows.

Ph.D. Advisor – This type of connection is defined to happen 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

Figure 2.4: 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 (2.1) and sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black). The blue dashed line shows the 45 degree line.

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.

The MLiM specification using the Publication Outcomes data is given by (2.1):

$$y_{i,r,t} = \alpha + \sum_{m=1}^{3} \sum_{j \neq i} w_{m,i,j,t} y_{j,r,t} \beta_m + \sum_{m=1}^{3} \sum_{j \neq i} w_{m,i,j,t} \widetilde{\mathbf{x}}_{j,r,t}^{\top} \boldsymbol{\delta}_m + \mathbf{x}_{i,r,t}^{\top} \boldsymbol{\gamma} + \lambda_r + \lambda_t + \lambda_0 + \varepsilon_{i,r,t}, \quad (2.15)$$

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 (2.15) 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 (2.15). 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 (2.15) 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 (2.15), we need to impose restrictions on the dependence patters on the population joint distribution $\mathcal{F}$. For this empirical application, I choose the constants K_c and K_d 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 (2.6) and (2.7). Table 2.1 presents the results for the Efficient GMM estimator characterized in Theorem 4. For comparison, the results for the first stage GMM estimator and the OLS estimator are presented in Tables 2.2 and 2.3 in Appendix 2.12. I conduct the analysis 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 2.1 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 4 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 2.1 and 2.2 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 4. Additionally, the OLS estimated peer and contextual effects parameters in Table 2.3 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 2.1 provides different meaningful results. First, as in Chapter 1 peer effects are found to be positive and statistically significant for articles' quality coming from the *Co-authors* network for all years. 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. The results for direct effects are all consistent with the results in Chapter 1. Finally, the empirical results in this Chapter also provide statistical evidence in favor of the exclusion restriction on the alumni network that I imposed in Chapter 1

2.8 Conclusion

This paper provides a novel approach to show how multilayer network data structures can be used to identify and consistently estimate network effects. My 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. I provide a simple linear GMM estimator and characterize its 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 me to construct correct standard errors for inference.

This method will be helpful to empirical researchers aiming to estimate network effects using observational data. My framework can handle both monolayer and multilayer data

Table 2.1: 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 ($\widehat{\boldsymbol{\gamma}}$)					
`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 (2.5.1) 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.

structures. However, it is more useful when the analyst observes at least two layers. The reason is that this method is designed to use all the additional information provided by including additional types of connections. 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).

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.

2.9 Appendix: 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 10 (a) and (b):

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

By Assumption 10 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 (2.7) 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 11, $\text{Cov}(x_{N,i,q}, \varepsilon_{N,i}) = \mathbb{E}(x_{N,i,q}, \varepsilon_{N,i})$, so that equations (2.6) and (2.6) 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 3. First note that Assumption 9 guarantees that the solution for model (2.1) exists. Fix K_c and K_d with $K_d > K_c + 1$, and let Assumption 10 and 11 holds such that Proposition 1 follows for any realization of $\mathcal{M}_N \in \mathscr{M}$. Assumption 12 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 $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]\,(\boldsymbol{\psi}^0 - \boldsymbol{\psi}) = 0$. Under the Assumption 13, it follows that $\mathbb{E}[\mathbf{m}(\boldsymbol{\psi})] = 0$ if and only if $\boldsymbol{\psi}^0 = \boldsymbol{\psi}$. □

Proof of Theorem 4. The GMM estimator is from Section 2.5 in the main text can be written as

$$\widehat{\boldsymbol{\psi}}_{GMM} = \boldsymbol{\psi} + (n^{-1}\mathbf{D}_n^\top \mathbf{Z}_n \left(\mathbf{A}_n^\top \mathbf{A}_n\right) n^{-1}\mathbf{Z}_n^\top \mathbf{D}_n)^{-1} n^{-1}\mathbf{D}_n^\top \mathbf{Z}_n (\mathbf{A}_n^\top \mathbf{A}_n) n^{-1}\mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n \tag{2.16}$$

By construction, the matrix $\mathbf{A}_n$ is assume to converge to $\mathbf{A}_N$, so that $(\mathbf{A}_n^\top \mathbf{A}_n) \to (\mathbf{A}_N^\top \mathbf{A}_N)$ as $n \to \infty$, which is assumed to be finite and full rank. From Lemma 10, $n^{-1}\mathbf{Z}_n^\top \mathbf{D}_n$ 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 13. Finally, Lemma 11 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 2.16

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

Let $\mathbf{Q}_{zx} = \mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top]$. From Lemmas 10 and 12 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 $\mathbf{A}_N = \boldsymbol{\Omega}_N^{-1/2}$ so that $\mathbf{A}_N^\top \mathbf{A}_N = \boldsymbol{\Omega}_N^{-1}$. With that choice of weighting matrix, the asymptotic variance-covariance matrix is given by

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

By the remarks presented in section 2.1 about conditional expectation interpretation of $\mathbf{Q}_{zx}$, it follows that $\boldsymbol{\Sigma}_N^\star$ can also be written as

$$\left[\left(\sum_{i\in\mathcal{I}_N}\mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top \mid \mathcal{H}^*_{N,\lambda,i},\mathcal{H}^*_{N,i}\neq\mathbf{O}_{R\times R}]\mathcal{K}_{N,i}\right)^\top \boldsymbol{\Omega}_N^{-1} \left(\sum_{i\in\mathcal{I}_N}\mathcal{K}_{N,\lambda,i}\mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top \mid \mathcal{H}^*_{N,\lambda,i},\mathcal{H}^*_{N,i}\neq\mathbf{O}_{R\times R}]\mathcal{K}_{N,i}\right)\right]^{-1}, \tag{2.17}$$

Equation (2.17) 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 $\widehat{\boldsymbol{\psi}}^\star_{GMM} = (\mathbf{D}_n^\top\mathbf{Z}_n\boldsymbol{\Omega}_n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n)^{-1}(\mathbf{D}_n^\top\mathbf{Z}_n\boldsymbol{\Omega}_n^{-1}\mathbf{Z}_n^\top\mathbf{y}_n)$. The previous arguments imply that $\sqrt{n}(\widehat{\boldsymbol{\psi}}^\star_{GMM}-\boldsymbol{\psi}) \xrightarrow{d} \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}^\star_N)$. Note that the consistency argument also applies to $\widehat{\boldsymbol{\psi}}^\star_{GMM}$ given that by Lemma 13, $\boldsymbol{\Omega}_n \to \boldsymbol{\Omega}_N < \infty$, as $n\to\infty$. □

2.10 Appendix: Proofs of Auxiliary Results

Lemma 6 (Invertibility of $\mathbf{S}(\boldsymbol{\beta})$)**.** *Let Assumption 9 holds then* $\mathbf{S}(\boldsymbol{\beta}) = \mathbf{I}_N - \sum_{m=1}^M \beta_m \mathbf{W}_{N,m}$ *is invertible.*

Proof. In the trivial case where $\beta_m = 0$ for all m, it follows that $\mathbf{S}(\beta) = \mathbf{I}_N$ 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 $\theta \neq 0$, $\mathbf{S}(\boldsymbol{\beta})$ can be written as

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

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 9, 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 9. Thus, $\Pi_i(1 - \theta\lambda_i) \neq 0$, completing the proof. □

Remark 2 (Series Expansion)**.** Assumption 9 also guarantees that it is possible to write the matrix $\mathbf{S}(\boldsymbol{\beta})$ as an infinite series given by $\mathbf{S}(\boldsymbol{\beta}) = \sum_{r=0}^{\infty} \left(\sum_{m=1}^{M} \beta_m \mathbf{W}_{N,m} \right)^r$. 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 (2.2) in the main text as an infinite sum of the product of different adjacency matrices given by

$$\mathbf{y}_N = \sum_{\zeta=0}^{\infty} \sum_{\phi\in\mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell\in\phi}\beta_\ell\right) \sum_{m=1}^{M} \mathbf{A}_N(\zeta,\phi)\mathbf{W}_{N,m}\mathbf{X}_N\boldsymbol{\delta} \\ + \sum_{\zeta=0}^{\infty} \sum_{\phi\in\mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell\in\phi}\beta_\ell\right) \mathbf{A}_N(\zeta,\phi)\widetilde{\mathbf{X}}_N\widetilde{\boldsymbol{\gamma}} + \sum_{\zeta=0}^{\infty} \sum_{\phi\in\mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell\in\phi}\beta_\ell\right) \mathbf{A}_N(\zeta,\phi)\boldsymbol{\varepsilon}_N, \tag{2.18}$$

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 $\mathbf{W}^k_{N,m}$ gives the number of paths of length k from agents i to j (for some layer m), see e.g., Graham (2015), 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, (2.18) 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. (2021), it is possible to rewrite (2.18) as

$$\begin{aligned}\mathbf{y}_N =& \alpha \sum_{r=0}^{\infty}\left(\sum_{m=1}^{M}\beta_m \mathbf{W}_{N,m}\right)^r \iota + \sum_{q=1}^{Q}\gamma_q \mathbf{x}_N^q + \sum_{q=1}^{Q}\sum_{m=1}^{M}\left(\sum_{r=0}^{\infty}\beta_m^r \mathbf{W}_{N,m}^{r+1}\right)\mathbf{x}_N^q\left(\delta_{q;m}+\gamma_q\beta_m\right)\\ &+\sum_{q=1}^{Q}\sum_{m\neq s\in\{1,\ldots M\}}\left(\sum_{r=0}^{\infty}\sum_{r'=0}^{\infty}\beta_m^r\beta_s^{r'}\mathbf{W}_m^{r+1}\mathbf{W}_s^{r'+1}\right)\mathbf{x}_N^q\left(\delta_{q;m}+\gamma_q\beta_m\right)+\ldots+\\ &+\sum_{r=0}^{\infty}\left(\sum_{m=1}^{M}\beta_m \mathbf{W}_m\right)^r \boldsymbol{\varepsilon}_N,\end{aligned} \tag{2.19}$$

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 $\max_{q\in\{1,\ldots,Q\}}\left(\left|\gamma_q\beta_1+\delta_{q;1}\right|,\ldots,\left|\gamma_q\beta_M+\delta_{q;M}\right|\right)\neq 0$.

Remark 3 (Solution for the Outcome Equation)**.** Equation (2.19) 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 6. 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 (2.18) can be written as

$$\begin{aligned}
\mathbf{y}_n = &\sum_{\zeta=0}^{\infty} \sum_{\phi \in \mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell \in \phi} \beta_\ell\right) \sum_{m=1}^{M} \mathbf{A}_n(\zeta, \phi)\mathbf{W}_{n,m}\mathbf{X}_n\boldsymbol{\delta} \\
&+ \sum_{\zeta=0}^{\infty} \sum_{\phi \in \mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell \in \phi} \beta_\ell\right) \mathbf{A}_n(\zeta, \phi)\widetilde{\mathbf{X}}_n\widetilde{\boldsymbol{\gamma}} + \sum_{\zeta=0}^{\infty} \sum_{\phi \in \mathbb{P}\left(\{1,\ldots,M\},\zeta\right)} \left(\prod_{\ell \in \phi} \beta_\ell\right) \mathbf{A}_n(\zeta, \phi)\boldsymbol{\varepsilon}_n.
\end{aligned} \tag{2.20}$$

Lemma 7. *Let Assumptions 10 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 14 hold for* $R_{i,j}$ *and* $R_{h,s}$. *Then*

$$|\mathrm{Cov}\,(R_{i,j}, R_{h,s})| \leq 2\bar{\theta}_{n,s}(C+16) \times 4\,(\pi_1 + \tilde{\gamma}_1)\,(\pi_2 + \tilde{\gamma}_2)\,\underline{\theta}_{n,s}^{1-p_f-p_g}, \tag{2.21}$$

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 10, 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 14 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 8 (LLN for Products of ψ-dependent Random Variables)**.** *Let Assumptions 10, 14, 15 and 16 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 I assume 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

$$\left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w_{i,j} \left(R_{n,i,j} - \mathbb{E}\left[R_{n,i,j} \right] \right) \right\|_1 \leq \left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w_{i,j} \left(R_{n,i,j} - \mathbb{E}\left[R_{n,i,j} \right] \right) \right\|_2 \tag{2.22}$$

where (2.22) 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

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

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

$$\begin{aligned} \operatorname{Var}\left(\sum_{j \in \mathcal{I}_i} w_{i,j} R_{n,i,j} \right) &= \sum_{j \in \mathcal{I}_i} w_{i,j}^2 \operatorname{Var}(R_{n,i,j}) + \sum_{j \neq s \in \mathcal{I}_i} w_{i,j} w_{i,s} \operatorname{Cov}(R_{n,i,j}, R_{n,i,s}) \\ &\leq a \sum_{j \in \mathcal{I}_i} w_{i,j}^2 + \sum_{j \in \mathcal{I}_i} \sum_{d \geq 1} \sum_{s \in \mathcal{P}_n(j,d) \cap \mathcal{I}_i} \left| \operatorname{Cov}(R_{n,i,j} R_{n,i,s}) \right| \\ &\leq a \sum_{j \in \mathcal{I}_i} w_{i,j}^2 + b \sum_{d \geq 1} \theta_{n,d} \sum_{j \in \mathcal{I}_i} \left| \mathcal{P}_n(j,d) \right|, \end{aligned} \tag{2.23}$$

where the second inequality follows form $w_{i,j}, w_{i,s} \in [0, 1]$, 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 10 and 14, and Lemma 7, 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 7, 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 (2.22), by the properties of the covariance, the expression inside the summation can be written as

$$\begin{aligned} \text{Cov}\left(\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}, \sum_{s\in\mathcal{I}_h} w_{h,s}R_{n,h,s}\right) &= \sum_{j\in\mathcal{I}_i}\sum_{s\in\mathcal{I}_h} w_{i,j}w_{h,s}\text{Cov}(R_{n,i,j}, R_{n,h,s}) \\ &\leq \sum_{j\in\mathcal{I}_i}\sum_{d\geq 1}\sum_{s\in\mathcal{P}_n(j,d)\cap\mathcal{I}_h} |\text{cov}(R_{n,i,j}R_{n,h,s})| \\ &\leq b\sum_{d\geq 1}\theta_{n,d}\sum_{j\in\mathcal{I}_i}|\mathcal{P}_n(j,d)|, \end{aligned} \tag{2.24}$$

where the inequalities in (2.24) follow from the same arguments discussed before. It follows from equations and (2.24) that the total variance can be bounded by

$$\begin{aligned} \text{Var}\left(\sum_{i\in\mathcal{I}_{n,m,\lambda}}\sum_{j\in\mathcal{I}_i} w_{i,j}R_{n,i,j}\right) &\leq a\sum_{j\in\mathcal{I}_i} w_{i,j}^2 + 2b\sum_{d\geq 1}\theta_{n,d}\sum_{j\in\mathcal{I}_i}|\mathcal{P}_n(j,d)| \\ &= a\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}^2 + 2b\sum_{d\geq 1}\theta_{n,d}\sum_{i\in\mathcal{I}_n}|\mathcal{P}_n(j,d)| \\ &\leq n\left(a\sum_{i\in\mathcal{I}_n} n^{-1}\sum_{j\in\mathcal{I}_i} w_{i,j} + 2b\sum_{d\geq 1}\theta_{n,d}\bar{D}_n(d)\right), \end{aligned} \tag{2.25}$$

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 2.22 and 2.25,

$$\left\| \frac{1}{n}\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w_{i,j}\left(R_{n,i,j} - \mathbb{E}\left[R_{n,i,j}\right]\right)\right\|_1 \leq \left(\frac{a}{n}\sum_{i\in\mathcal{I}_n}\bar{\mathcal{I}}_i^w + \frac{2b}{n}\sum_{d\geq 1}\theta_{n,d}\bar{D}_n(d)\right)^{1/2} \tag{2.26}$$

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 15 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 16 guarantees that $n^{-1}\sum_{d\geq 1}\theta_{n,d}\bar{D}_n(d) \xrightarrow{\text{a.s.}} 0$. It follows that the right hand side of equation (2.26) is $o_{a.s.}(1)$, completing the proof. □

Lemma 9 (LLN for Outcomes and Regressors)**.** *Let Assumptions 9, 10, 14, 15 and 16 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,*

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\mathbf{w}_{n,m,\lambda,i}\mathbf{x}_{n,q}\mathbf{w}_{n,s,i}\mathbf{y}_n - \mathbb{E}[\mathbf{w}_{N,m,\lambda,i}\mathbf{x}_{N,q}\mathbf{w}_{N,s,i}\mathbf{y}]\right\|_1 \xrightarrow{a.s.} 0 \tag{2.27}$$

and

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\mathbf{w}_{n,m,i}\mathbf{y}_n\mathbf{x}_{n,q} - \mathbb{E}[\mathbf{w}_{N,m,i}\mathbf{y}x_{N,q,i}]\right\|_1 \xrightarrow{a.s.} 0 \tag{2.28}$$

Proof. From Assumption 9, 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:

$$\begin{aligned}\left(\prod_{\ell\in\phi}\beta_\ell\right)&(\mathcal{W}_{n,m_1,\lambda}\mathbf{x}_{n,q})^\top\mathbf{W}_{n,m_2}\mathbf{A}(\zeta,\phi)\mathbf{W}_{n,m_3}\mathbf{x}_{n,l}\delta_{m_3,l}\\ &=\left(\prod_{\ell\in\phi}\beta_\ell\right)(\mathcal{W}_{n,m_1,\lambda}\mathbf{x}_{n,q})^\top\mathbf{A}(\zeta+2,\{m_2,\phi,m_3\})\mathbf{x}_{n,l}\delta_{m_3,l},\end{aligned}\tag{2.29}$$

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 rigth hand side of equation (2.29) can be written as

$$\frac{1}{n}\sum_{\ell\in\mathcal{I}_n}\sum_{i,j\in\mathcal{I}_\ell}w_{i,j}R_{n,i,j},\tag{2.30}$$

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 8, it follows that (2.30) 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 9. This completes the proof for the convergence of the first component of the outcome equation $\mathbf{y}_n$ in (2.20).

For the second component in (2.20), 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 (2.20) 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 8 also applies for this case. This completes the proof for equation (2.27). The proof for (2.28) is analogous. □

Lemma 10 (LLN for Instruments and Regressors)**.** *Let Assumptions 10, 14, 15 and 16 hold. Then as* $n\to\infty$

$$\left\| \frac{1}{n}\sum_{i\in\mathcal{I}_n}\mathbf{z}_{n,i}\mathbf{d}_{n,i}^\top - \mathbb{E}[\mathbf{z}_{N,i}\mathbf{d}_{N,i}^\top] \right\|_1 \xrightarrow{a.s.} 0. \tag{2.31}$$

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 (2.31), it suffices to show the convergence for each of the four components described before. Firs, from Lemma 9, 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 8. 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 8 component-wise for $\sum_{i\in\mathcal{I}_n} \mathbf{z}_{n,i}\mathbf{d}_{n,i}^\top$ implies the result. □

Lemma 11 (LLN for Instruments and Errors)**.** *Let Assumptions 10, 14, 15 and 16 hold. Then as $n \to \infty$*

$$\left\| \frac{1}{n} \sum_{i\in\mathcal{I}_n} \mathbf{z}_{n,i}\varepsilon_{n,i}^\top - \mathbb{E}[\mathbf{z}_{N,i}\varepsilon_{N,i}^\top] \right\|_1 \xrightarrow{a.s.} 0. \tag{2.32}$$

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 10 parts (1) and (3). □

Lemma 12 (Central Limit Theorem)**.** *Let Assumptions 10, and 14 to 17 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 $n \to \infty$*

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

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 7, 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 13 (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 10, and 14 to 16 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 7 combined with the arguments in Lemma 8 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 (2.25), 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 10), 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 14 (Multivariate Central Limit Theorem)**.** *Let Assumption 10, and 14 to 17 hold. Then,*

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

Proof. From Lemma 12, $n^{-1/2}\sum_{i=1}^n z_{n,i,q}\varepsilon_{n,i} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_n^2)$. From Lemma 13, $\boldsymbol{\Omega}_N$ exists. Therefore, the result follows from the Cramér-Wold device. □

2.11 Appendix: Multilayer Shortest Path Algorithms

This section describes the computation process required to calculate the empirical analog of the moment condition matrices. 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 $E \equiv \bigcup_{m \in M} E_m \bigcup \mathcal{C}$ (Balasubramanian et al., 2022). 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 2.7 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 (Balasubramanian et al., 2022). 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 2.6 and 2.7.

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 $s \in \bigcup_{m \in M} V_m$.

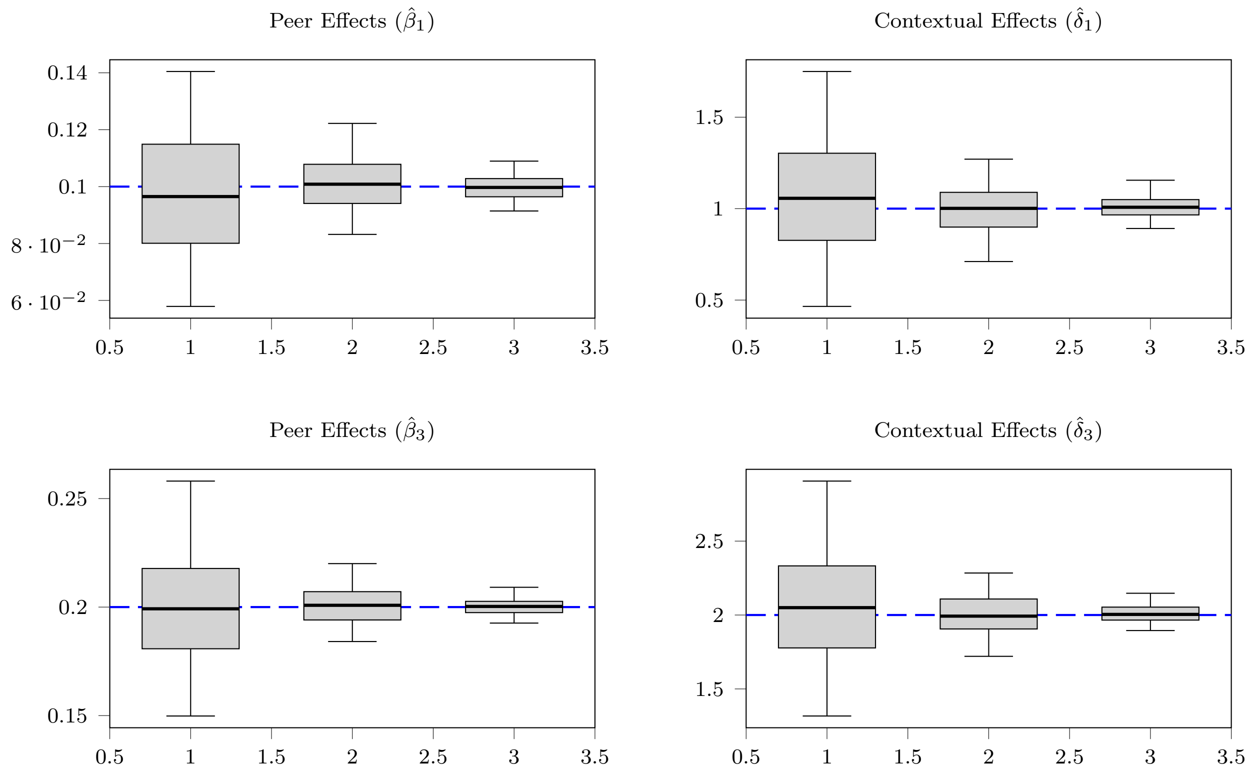
Output: second shortest paths and color changes for all nodes $v \neq s$ in $\mathcal{M}$.

```
foreach v s ∈ ⋃_{m∈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;
```

2.12 Appendix: Additional Simulation and Estimation Results

This section contains additional simulation and estimation results. Plots 2.5 and 2.6 show the results from the Monte Carlo simulation for the additional network effects $[\beta_1^0, \beta_3^0, \delta_1^0, \delta_3^0]$.

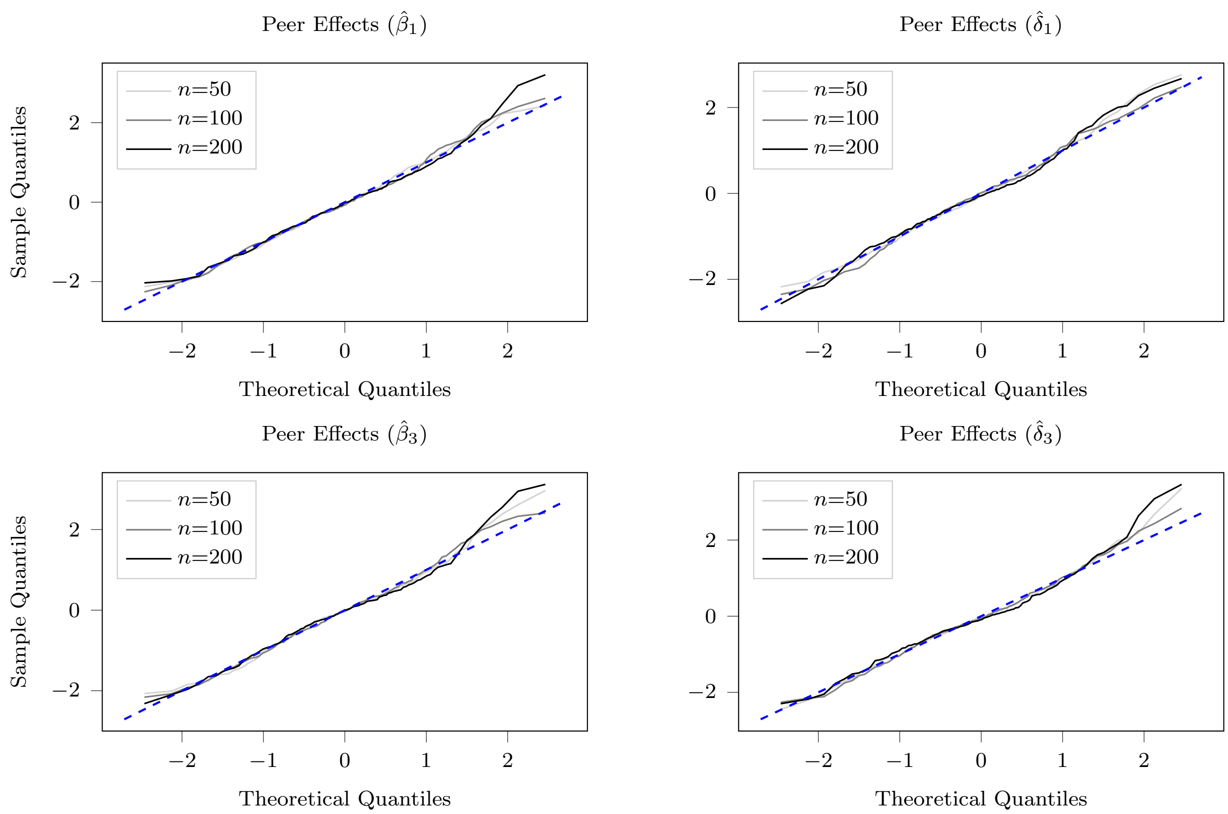
Figure 2.5: Box Plots for the GMM Estimator



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.

The results for the other parameters are presented in Section 2.6 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 2.2 and 2.3 present the empirical estimation results using the publications data described in section 2.7. These results are used as a benchmark to compare the behavior of the efficient GMM estimator.

Figure 2.6: 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 (2.1) and sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black). The blue dashed line shows the 45 degree line.

Table 2.2: 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 ($\widehat{\boldsymbol{\gamma}}$)					
`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 (2.5.1) 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 2.3: 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 (2.5.1) 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.

Chapter 3

Inference in Network Formation Models with Payoff Externalities

Note: I want to thank my colleagues Cheng Ding and Santiago Montoya-Blandón for their comments and support. This chapter would not exist without your help.

This paper provides a novel approach to identify and perform inference on the utility parameters of a network formation model with payoff externalities using observed network data. The existence of externalities induces an issue of multiple equilibria. Under the assumption that a predetermined probability distribution exists over the set of all possible equilibrium networks, we show that local point identification of the parameters of interest is possible. We propose a Bayesian estimation method to conduct statistical inference of the structural payoff coefficients. We address the issue of high-dimensional numerical integration by proposing a composite likelihood function based on the marginal distribution of all the possible subgraphs forming the observed network. We present an empirical application to model the network

formation process of individuals creating social connections in villages in Karnataka, India. We find strong evidence of homophily effects.

3.1 Introduction

Social networks are critical to determining how individuals make choices in contexts ranging from labor markets and educational achievement to substance abuse and criminality (Gaviria and Raphael, 2001; Sacerdote, 2001b; Bayer et al., 2009; Mas and Moretti, 2009). As network structures arise from individual strategic interaction, it becomes essential to understand the underlying network formation process when studying the effects of these structures using observational data (see, e.g., Goldsmith-Pinkham and Imbens, 2013; Johnsson and Moon, 2021). The main contribution of this article is to propose a method to point-identify and estimate strategic network formation models using observational data on network links. In particular, we offer a Bayesian approach to estimate the structural parameters of preferences and equilibrium selection probabilities that characterize the network formation model. For our method, observed networks are interpreted as the equilibrium outcome of a complete information game where individuals make connections based on a flexible payoff function that allows for externalities in the utility received from links elsewhere in the network (Pelican and Graham, 2020).

Allowing for strategic externalities is appealing since it provides a rich model that matches documented features of social networks found in the data, such as clustering and homophily (Jackson and Rogers, 2007; Jackson et al., 2012; Sheng, 2020). However, strategic network formation models with utility externalities are plagued with identification and estimation issues. These outstanding problems include multiple equilibria, statistical dependence in large networks, and the curse of dimensionality (Leung, 2015; de Paula et al., 2018). Differing from previous research proposing methods that resolve some issues but not all, this paper provides a technique that simultaneously addresses all these fundamental issues.

We deal with the problem of multiple equilibria by specifying both a payoff function and an equilibrium selection mechanism as the primitives of the network formation model. The selection mechanism determines the probability distribution over possible equilibria for a given realization of exogenous characteristics, individual heterogeneity, and idiosyncratic dyadic shocks. As in Bajari et al. (2010), we take an empirical approach to characterize the equilibrium selection. Critically, we assume the existence of a predetermined and potentially unknown probability distribution over all possible equilibrium outcomes. We interpret the probability distribution as nature assigning different likelihoods of occurrence to different network structures. We show that this assumption is enough to guarantee local point identification of the structural parameters underlying both utility functions and equilibrium selection probabilities. Our approach differs from previous studies of discrete games of complete information with multiple equilibria such as Bajari et al. (2010) and Bajari et al. (2011) due to the added complexity in the strategy space for the game we consider. This complexity makes it intractable to compute all possible equilibria of the game. Therefore, instead of defining a selection mechanism in the space of individuals' strategies, this paper focuses on selecting network structures that are the game's equilibrium outcomes. One advantage of directly selecting outcomes is that selection of equilibrium networks can result from both pure and mixed strategies (Pelican and Graham, 2020).

There has been an increasing interest in proposing identification and estimation methods to deal with the issues of multiple equilibria and the curse of dimensionality. Significant developments in this literature include modeling the network formation as a sequential process (Mele, 2017; Christakis et al., 2020), focusing only on non-negative externalities (Miyauchi, 2016), and using subnetworks as the unit of analysis (Chandrasekhar, 2016; Sheng, 2020). To the best of our knowledge, our paper contributes to this literature by providing the first results on point-identification of both preference and selection mechanisms parameters. In addition, we also offer a method to conduct statistical inference on those parameters that do not rely on asymptotic approximations. This contribution is relevant because it can handle large and repeated network data structures.

We then turn to the issue of estimating the identified quantities. To this end, we introduce some assumptions on equilibrium selection based on the idea of *multiplicity regions*. We define a multiplicity region as a set in the space of idiosyncratic dyadic shocks where the group of all equilibrium networks is the same for any value of the shocks belonging to that region. We show that the complete likelihood of observing a network $\mathbf{d}$ can be written as a weighted sum of integrated likelihoods across all possible multiplicity regions. Though critical for showing that the model's parameters are point identified, the formalization in terms of multiplicity regions does not solve the curse of dimensionality problem inherent to the estimation process. This is due to the dimensionality of the idiosyncratic shocks, which result in a computationally intractable likelihood, even when using simulation-based methods (McFadden, 1989; Hajivassiliou and McFadden, 1998).

To overcome this issue, we resort to a Bayesian estimation method. In particular, we propose a Markov Chain Monte Carlo (MCMC) algorithm similar in spirit to Mele (2017). Following a data augmentation approach, we introduce the idiosyncratic shocks as latent variables to be sampled along with the necessary preference parameters. This idea forms the basis of many computationally efficient algorithms for estimating models where numerical integration is unfeasible, such as discrete choice models with large choice sets (Albert and Chib, 1993; McCulloch et al., 2000). With this choice of latent variables, our algorithm alternates between sampling from the utility parameters conditional on a transformation of the shocks and the distribution of shocks given parameters and network data.

The advantages of using a Bayesian estimation method are twofold. First, it allows us to sidestep the issue of high-dimensional numerical integration to evaluate the likelihood. Second, we can address the potential issue of statistical dependence in large networks. The key idea is that because our estimation method does not rely on asymptotic approximations to conduct inference, distinguishing between large-games and many-games asymptotics becomes less relevant. Notably, the flexibility in our estimation method can accommodate both large-networks and many-networks data structures. Finally, our Bayesian approach also generates posterior distributions for the equilibrium selection probabilities, which allows the researcher to simulate the model and perform counterfactual analysis.

To highlight the usability of our approach, we present an empirical application using data from the Social Networks and Microfinance project, which contains the publicly available data of participation in a program of Bharatha Swamukti Samsthe (BSS), a microfinance institution (MFI) in rural southern Karnataka (Banerjee et al., 2013). The data include information on thirteen possible relationships among individuals, including visiting each other, praying, borrowing and lending money and goods, obtaining advice, and giving advice. We combine the different social ties into one unique social network encompassing the thirteen dimensions. Based on the constructed social network, we model the network formation process and estimate the payoff parameters of interest. We find strong evidence of homophily for most of the characteristics in our analysis. The most substantial homophily effects happen among the same gender and working status of villagers.

The structure of the paper is as follows. Section 3.2 introduces the network formation model, our population assumptions, and the main identification result. Section 3.3 presents the Bayesian algorithm, and introduces the idea of the composite likelihood function. Section 3.4 presents the data and empirical results. Finally, section 3.5 concludes.

Notation: we follow the standard statistical notation and use capital letters to describe random variables and lower case letters represent the realizations of those random variables. We use bold letters for vectors and non-bold letters for scalars. For example, $\mathbf{X}$ is a random vector while $\mathbf{x}$ is a vector of realizations, and U is a scalar random variable with u is a possible realization.

3.2 Network Formation Model and Identification

3.2.1 Network Formation

Following Pelican and Graham (2020) and Sheng (2020), we assume that observed network structures are the result of a network formation model where N individuals choose connections simultaneously in what is known as a link announcement game. Each individual i is characterized by a vector of observed attributes X_i, and the unobserved (to the researcher) shocks A_i and U_{ij} for $j \neq i$. Let $\mathbf{d} \in \mathbb{D}_N$ be a realization of a network configuration from the set $\mathbb{D}_N$ of all possible N-player equilibrium networks. We assume payoff functions that allow players to have preferences over other individuals' positions in the network. In particular, we assume that individuals have additively separable utility functions divided into degree heterogeneity, assortative matching (also known as homophily) and an externality component that can generate multiple equilibria. In defining our payoff function, for notational convenience, we do not explicitly include $\mathbf{A}$ and $\mathbf{X}$ as arguments.

Furthermore, for estimation purposes, we will introduce a few mild assumptions on the distribution of $\mathbf{A}$ given the covariates $\mathbf{X}$. These assumptions will resemble a correlated random effects (CRE) approach that differs from that in Graham (2017), which instead resembles a fixed-effects type assumption.[1] In fact, combined with our Bayesian estimation method outlined in Section 3.3, this view offers an alternative to Graham (2017) for estimation and testing when we assume that there are no additional payoff externalities. Individual's utility from the network configuration is determined by

$$\nu_i\left(\mathbf{d}, U_{ij}\right) = \sum_j d_{ij}\left[A_i + A_j + X_i' \Lambda_0 X_j + \gamma_0 s_{ij}(\mathbf{d}) - U_{ij}\right], \quad i = 1, \ldots, N \tag{3.1}$$

where the dependence on the network $\mathbf{d}$ follows from the fact that individual i's utility can be potentially affected by other individuals' links in the network. The utility function depends on two individual degree heterogeneity components A_i and A_j, a dyadic unobserved component U_{ij}, a homophily component $X_i' \Lambda_0 X_j$, and preferences over links other than d_{ij} given by the function $s_{ij}(\mathbf{d})$. The function $s_{ij}(\mathbf{d})$ can include relevant cases in the literature such as taste for reciprocated links in directed networks (Pelican and Graham, 2020), and

[1] We use the language of panel data models to make this distinction. In this literature, a fixed effect is usually an individual-specific effect correlated to covariates. Still, it can be eliminated from the model using a particular transformation (e.g., differencing in linear panel models). Instead, assumptions in place in a CRE framework restrict the distribution of effects given covariates but generally make the model more tractable.

taste for indirect connections and completing triangles in undirected networks (Mele, 2017; Christakis et al., 2020; Sheng, 2020).

As pointed out by Pelican and Graham (2020), the selection of an equilibrium selection concept is closely related to the assumption on the directed or undirected nature of individuals' links. Games on directed networks are associated with the Nash equilibrium solution concept, while in the analysis of undirected networks, pairwise stability as equilibrium concept plays a fundamental role (Jackson and Wolinsky, 1996). To accommodate the structure of our empirical application, we assume that individuals form undirected connections. In 3.7, we show that this assumption is not necessary for our identification and estimation results as they also apply to a game where individuals can form directed links under a Nash equilibrium solution concept. Following the literature on undirected network formation games, we use the concept of pairwise stability under non-transferable utility as our equilibrium concept. Individuals i and j decide whether or not to form a connection based on their marginal utilities. From the payoff function (3.1), it follows that the marginal utility of creating a link between i and j is

$$\Delta\nu_{ij}\left(\mathbf{d}_{-ij}; U_{ij}\right) = A_i + A_j + W_{ij}'\lambda_0 + \gamma_0 s_{ij}(\mathbf{d}) - U_{ij}\,, \quad i = 1, \ldots, N; j = 1, \ldots, N; i \neq j\,, \tag{3.2}$$

where $\mathbf{d}_{-ij}$ represent the set of all connections in network $\mathbf{d}$ excluding the ij-link, $W_{ij} = X_i \otimes X_j$ and $\lambda = \mathbf{\Lambda'}$. Using this definition of marginal utility in equation (3.2), we can now introduce the definition of pairwise stability.

Definition 4 (Pairwise Stability)**.** *A network* $\mathbf{d}$ *is pairwise stable under non-transferable* *utility if*

(i) for any $d_{ij} = 1, \Delta\nu_{ij}(\mathbf{d}_{-ij}; u_{ij}) \geq 0$ *and* $\Delta\nu_{ji}(\mathbf{d}_{-ji}; u_{ji}) \geq 0$*;*

(ii) for any $d_{ij} = 0, \Delta\nu_{ij}(\mathbf{d}_{-ij}; u_{ij}) > 0$ *implies* $\Delta\nu_{ji}(\mathbf{d}_{-ji}; u_{ji}) < 0$*.*

Equilibrium. We assume a game of complete information. Each individual i observes $\{A_i, X_i'\}_{i=1}^N$ and $\{u_{ij}\}_{i \neq j}$, then decides a set of links from the $N-1$ agents. A link is formed if both individuals perceive a positive marginal utility from the connection. Sheng (2020) shows that under non-transferable utility, assuming that the links are strategic complements, the model can be casted into a supermodular game in which the the existence of an equilibrium follows from the fixed-point theorem for isotone mappings (Topkis, 1979; Milgrom and Roberts, 1990). Therefore, assuming that $\gamma_0 \geq 0$ in our payoff function guarantees a non-empty set of pairwise stable networks $\mathbb{D}_N$.

3.2.2 Population Assumptions

We assume that the population is formed by an arbitrarily large set of individuals $\mathcal{I}_N$. An equilibrium population network $\mathbf{d} \in \mathbb{D}_N$ connects all nodes in $\mathcal{I}_N$. As argued by Goldsmith-Pinkham and Imbens (2013), the ability to identify the statistical properties of the network rests on the assumptions on the individuals' dependence on the population given the network structure. Two standard assumptions have been used in the econometrics literature of network formation games. The more straightforward but less realistic case assumes a large number of exchangeable networks. This assumption is consistent with a large population that can be divided into independent networks that are mutually disjoint. A more realistic alternative is to assume a large network defined on the set of nodes $\mathcal{I}_N$. As argued by Leung

(2015), the issue with this approach is that the researcher has to impose conditions on the network dependence between individuals to be able to use asymptotic approximations to conduct inference.

One of our Bayesian estimation method's key advantages is that we do not require an asymptotic approximation to conduct inference. Therefore, our approach can accommodate both of the mentioned population assumptions. However, to identify the parameters of the equilibrium selection mechanism, we require observing repeated networks. The identification argument applies to the repeated networks population assumption. However, in the case of the large networks, we have to assume further that the analyst can partition the large network into approximately independent sub-networks, see, e.g., Sheng (2020). It is then possible to estimate selection probabilities when multiple equilibria are present with repeated sub-network sequences.

3.2.3 Identification

The network externality component $s_{ij}(\mathbf{d})$ in our network formation game induces the potential for multiple equilibria. To specify a likelihood function for this problem, we need to incorporate a way to assign probabilities to equilibrium outcomes realized from the set of all possible equilibria. To that end, we define $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) : \mathbb{D}_N \times \mathbb{R}^N \to [0, 1]$ to be a function that assigns probabilities to the set of equilibrium networks, where $\theta = [\lambda', \gamma]'$ is the vector of payoff parameters. We call this function $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ the equilibrium selection function.[2] The set of networks in equilibrium can include both outcomes that result from individuals

[2]For convenience, we do not make explicit that the equilibrium selection distribution also depends on covariates $\mathbf{X}$ and degree heterogeneity effects $\mathbf{A}$ and $\mathbf{B}$.

playing mixed strategies or multiple equilibrium outcomes where individuals are playing pure strategies. With this definition, the likelihood of observing a network $\mathbf{d}$ is given by

$$P(\mathbf{d};\theta) = \int_{\mathbf{u}\in\mathbb{R}^N} \mathcal{N}(\mathbf{d},\mathbf{u};\theta) \prod_{i\neq j} f_U(u_{ij})\mathbf{u}\,. \tag{3.3}$$

For now, we leave f_U unspecified but we note that an usual assumption is the logistic distribution given by $f_U(u) = e^u / [1+e^u]^2$. This distributional assumption generates an exponential random graph structure that is common in network formation models. In our Bayesian algorithm, we replace this with a normality assumption that is more compatible with posterior updating.

A common practice in the econometrics of games in which multiple solutions are possible is to partition the errors' space into different subspaces that determine the potential number of equilibria conditional on given values of the regressors and the parameters of the model (De Paula, 2013). Characterizing these regions in our context is relevant because it allows us to express the likelihood of observing any given outcome as a sum of the probability mass in the areas where the outcome can happen weighted by the relative probability of the outcome in those regions. We call the areas partitioning the errors' space the *multiplicity regions*. In this paper, the definition of multiplicity regions requires conditioning on the values of regressors $\mathbf{X}$, degree heterogeneity $\mathbf{A}, \mathbf{B}$ and parameters θ.

Definition 5 (Multiplicity Region)**.** *Conditional on values of* $\mathbf{X}$*,* $\mathbf{A}$*,* $\mathbf{B}$ *and* θ*, a multiplicity region* $m \in M$ *in the space of* $\mathbf{U}$ *is a region where the set of all possible equilibrium networks is the same for all* $\mathbf{u} \in m$*. That is,* $\mathcal{N}(\mathbf{d},\mathbf{u};\theta) = \mathcal{N}(\mathbf{d},\mathbf{u}';\theta)$ *for any* $d \in \mathbb{D}_N$ *and* $\mathbf{u},\mathbf{u}' \in m$*. Furthermore, multiplicity regions partition the space of* $\mathbf{U}$*; i.e., for* $m, s \in M$*,* $m \cap s = \varnothing$

(regions are disjoint) and $\bigcup_{m \in M} m = \mathbb{R}^{N(N-1)}$.

A simple example can clarify the role of multiplicity regions in determining the observed network $\mathbf{d}$.

Example 3.2.1. *Consider the simple case when* $A_i = 0, B_j = 0$ *for all* i, j *and* $\Lambda = 0$. *Moreover, assume that externalities are given by intransitive triads,* $s_{ij}(\mathbf{d}) = d_{ik}d_{kj}$ *for all* i*, for* $\gamma \geq 0$. *Finally, assume* $N = 3$. *In this scenario there are only eight possible network configurations given by*

$$d_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, d_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, d_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, d_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, d_5 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$d_6 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, d_7 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \text{ and } d_8 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Determining what network configuration would emerge and whether or not it is unique, depends on the realizations of the errors and the values of the parameters. In this example the space of the vector of shocks $\mathbf{u} = (u_{12}, u_{21}, u_{13}, u_{31}, u_{23}, u_{32})$ *can be divided into the following partition that completely determine the admissible equilibrium networks given the externality parameter* γ*:*

Table 3.1: Possible Equilibrium Networks in a Simple 2x2 Example

(u_{12}, u_{21})	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
I_1, I_1	0	0	0	0	0	0	0	1
I_1, I_2	0	0	1	0	0	0	0	1
I_1, I_3	0	0	1	0	0	0	0	0
I_2, I_1	0	0	1	0	0	0	0	1
I_2, I_2	0	0	1	0	0	0	0	1
I_2, I_3	0	0	1	0	0	0	0	0
I_3, I_1	0	0	1	0	0	0	0	0
I_3, I_2	0	0	1	0	0	0	0	0
I_3, I_3	0	0	1	0	0	0	0	0

Note: table with all the possible multiplicity regions for an example with intransitive triads externalities. Each row represents a possible combination of realized shocks that follow in one of three possible buckets as described in equation (3.4). Each column represents the network structures that are possible equilibria. Position i, j in the table equals one if the network configuration j is a possible equilibrium in the multiplicity region i, and zero otherwise.

$$\underbrace{(-\infty, 0]}_{I_1} \cup \underbrace{(0, \gamma]}_{I_2} \cup \underbrace{(\gamma, \infty)}_{I_3}. \tag{3.4}$$

Table 3.1 presents all the possible equilibrium networks given that $u_{13}, u_{31} \in I_1$ and $u_{23}, u_{32} \in I_2$. Each row represents a realization of the errors such that each dyad shock falls into one of the three possible multiplicity regions. Intuitively, if the shocks are too small (both u_{12} and u_{21} fall into I_1), then it is always profitable for individuals to form connections, and the only equilibrium network is d_4. The same intuition holds for large values of the shocks where the only possible equilibrium configuration is the empty network. When both shocks fall into intermediate values (I_2), multiple equilibria arises, both the empty and the complete network are possible outcomes.

The primary identification idea is to separate the likelihood of the problem into all possible multiplicity regions and evaluate whether it is possible to identify the parameters of interest for each of those regions. Identification in this context is in terms of observational equivalence; i.e., data distribution at the true parameter is different from that at any other possible parameter value. We impose assumptions on equilibrium selection in terms of multiplicity regions sufficient to point-identify utility function parameters and selection probabilities from observational data on links. We separate $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta)$ into two components. One depends only on the network and multiplicity regions, while the other depends only on the network and errors.

Assumption 18 (Equilibrium Selection Characterization)**.** *Let $h_0(\mathbf{d})$ be the true predetermined probability distribution over the set of possible equilibrium networks $\mathbb{D}_N$.*

The most critical part of Assumption 18 is the existence of a predetermined probability distribution over the set of all possible equilibrium outcomes. Let $\mathbb{D}_m$ be the set of all possible equilibrium networks within an arbitrary multiplicity region $m \in M$. Then, from the definition of conditional probability, the true equilibrium selection distribution is such that

$$\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta_0) = \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{d' \in \mathbb{D}_m} h_0(\mathbf{d}')} \mathbb{I}\{\mathbf{u} \in m\} g(\mathbf{d}, \mathbf{u}; \theta_0).$$

Therefore, Assumption 18 implies that the selection probability $\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta_0)$ only depends on $\mathbf{X}$ and utility components via the multiplicity regions. We interpret this probability distribution as nature determining what type of equilibrium is more likely given some parameters and shocks. Because we are providing a non-parametric form for the selection probabilities, we partially alleviate the potential misspecification issues raised in De Paula (2013). A potential misspecification problem in our context would arise if the probability distribution $h_0(\mathbf{d})$ is indeed a function of any of the random variables $\mathbf{X}$, $\mathbf{A}$, $\mathbf{B}$ or $\mathbf{U}$. Accordingly, the second part of Assumption 18 explicitly assumes that the first component of the equilibrium selection distribution is independent of idiosyncratic shocks, the regressors, and the degree heterogeneities. Finally, the equilibrium selection probability depends on $\mathbf{u}$ because those shocks control the potential configurations that the network can take given a set of parameters. In particular, the function $g(\mathbf{d}, \mathbf{u}; \theta)$ takes the form

$$g(\mathbf{d}, \mathbf{u}; \theta) = \prod_{i=1}^{N} \prod_{j>i} \left[\mathbb{I}\left(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}) \geq u_{ij}\right) \mathbb{I}\left(A_j + B_i + W_{ji}'\lambda + \gamma s_{ji}(\mathbf{d}) \geq u_{ji}\right)\right]^{d_{ij}}$$
$$\times \left[1 - \mathbb{I}\left(A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}) \geq u_{ij}\right) \mathbb{I}\left(A_j + B_i + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}) \geq u_{ji}\right)\right]^{1-d_{ij}}. \quad (3.5)$$

Combining equation (3.3) with the definition of the equilibrium selection distribution in Assumption 18, it follows that the likelihood of the problem can be written as

$$\begin{aligned} P(\mathbf{d}; \theta, h_0) &= \int_{\mathbf{u} \in \mathbb{R}^N} \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{d' \in \mathbb{D}_m} h_0(\mathbf{d}')} \mathbb{I}\{\mathbf{u} \in m\} g(\mathbf{d}, \mathbf{u}; \theta) \prod_{i=1}^{N} \prod_{j \neq i} f_U\left(u_{ij}\right) \mathbf{u} \\ &= \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{d' \in \mathbb{D}_m} h_0(\mathbf{d}')} \int_{\mathbf{u} \in m} g(\mathbf{d}, \mathbf{u}; \theta) \prod_{i=1}^{N} \prod_{j \neq i} f_U\left(u_{ij}\right) \mathbf{u} \end{aligned} \quad (3.6)$$

Notice that when the network $\mathbf{d}$ is not an equilibrium in the multiplicity region m and the shocks $\mathbf{u}$ belong to that multiplicity region, the function $g(\mathbf{d}, \mathbf{u}; \theta)$ has to take the value of zero. The reason is that, given our definition of multiplicity region, if $\mathbf{u} \in m$ and $\mathbf{d}$ is not an equilibrium in m, then there exist a dyad ij such that it cannot be the case that $d_{ij} = 1$ and $A_i + B_j + W_{ij}'\lambda + \gamma s_{ij}(\mathbf{d}) \geq u_{ij}$ at the same time or *vice versa*. For simplicity in notation, let $G(\mathbf{d}, m; \theta) = \int_{\mathbf{u} \in m} g(\mathbf{d}, \mathbf{u}; \theta) \prod_{i=1}^{N} \prod_{j \neq i} f_U(u_{ij}) \, \mathbf{u}$. It follows that the likelihood can be represented as

$$P(\mathbf{d}; \theta, h_0) = \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{d' \in \mathbb{D}_m} h_0(\mathbf{d}')} G(\mathbf{d}, m; \theta). \tag{3.7}$$

The likelihood in (3.7) varies as a function of θ from the selection probability, the density of $\mathbf{u}$, and the shape of multiplicity regions. When testing for the existing of strategic interactions, Pelican and Graham (2020) define the concept of bucket, which is a mapping from the space of $\mathbf{u}$ to a collection of intervals that determine the connection behavior of individuals i and j. In particular, Pelican and Graham (2020) argue that if the realization of the shock u_{ij} falls into the outer buckets, then ij's connection decision is uniquely determined, while u_{ij} falling into the inner buckets opens the possibility for multiple equilibria. Following Pelican and Graham (2020), we define buckets as follows.

Definition 6 (Buckets)**.** *Let $s_{ij}(\mathbf{d})$ be an arbitrary type of network externality such that $s_{ij}(\mathbf{d}) \in \{\underline{s}, \dots, \bar{s}\}$, where $\underline{s}$ and $\bar{s}$ are the minimum and maximum values that s_{ij} can take. Then, an outer bucket is either the interval $(-\infty, \mu_{ij} + \gamma \underline{s}]$ or $(\mu_{ij} + \gamma \bar{s}, \infty)$, while the inner buckets are $(\mu_{ij} + \gamma s_k, \mu_{ij} + \gamma s_{k+1}]$ for all $\underline{s} \leq s_k < \bar{s}$.*

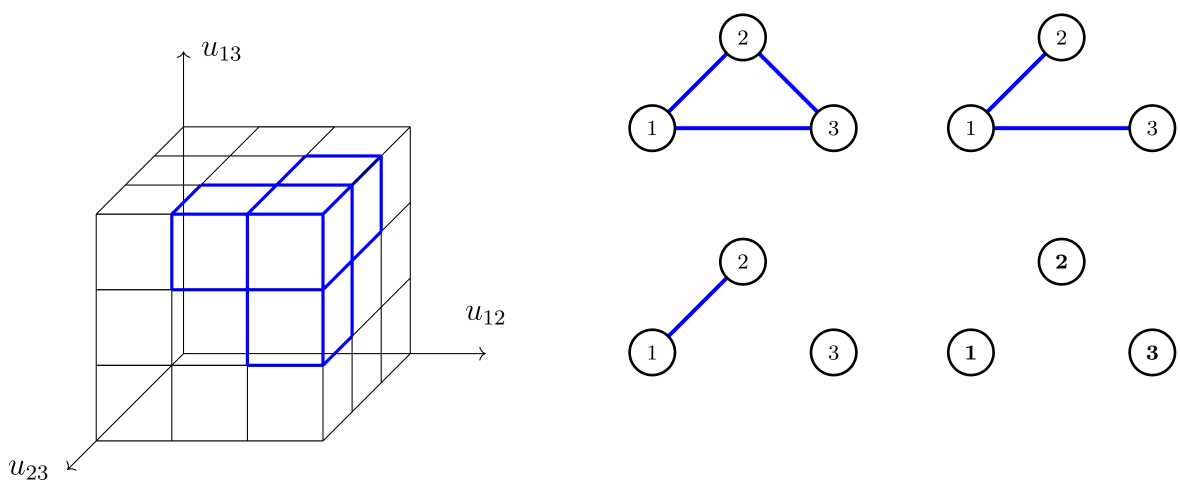
Definition 6 states that the externality function is discrete and takes on finitely many values between $\underline{s}$ and $\bar{s}$. For example, when s_{ij} takes the form of the reciprocity externalities, then $\underline{s} = 0$ and $\bar{s} = 1$, while if s_{ij} is the intransitive triads externality function, then $\underline{s} = 0$ and $\bar{s} = N - 2$. Example 3.2.1 presents a simple case where the definition of $s_{ij}(\mathbf{d})$ allows for a simple characterization of the multiplicity regions. In that example, the multiplicity regions coincide with the buckets as defined by Pelican and Graham (2020). However, for a more general externality measure such as intransitive triads, in which $s_{ij}(\mathbf{d}) = \sum_k d_{ik}d_{kj}$, the multiplicity regions will not coincide with the buckets. Even though the two measures do not coincide, there is a tight relationship between them. All multiplicity regions are composed of a countable number of buckets that also determine their shape. The following Proposition formalizes this discussion.

Proposition 2 (Multiplicity Regions Characterization)**.** *Let $b \in \mathbb{B}$ be an arbitrary bucket, where $\mathbb{B}$ is the set of buckets that partition the errors' space $\mathbb{R}^N$. Then, b must be within one and only one multiplicity region $m \in M$, where $b \subseteq m$.*

The proof of proposition 2 is relegated to 3.6. Instead, we provide an example with the main intuition of the result. Example 3.2.2 shows the geometric intuition from the result in proposition 2 for a simple case of three individuals when the taste for completing triangles generates the externalities.

Example 3.2.2. *Assume $\mathcal{I}_3 = \{1, 2, 3\}$ with utility functions given by equation (3.1), where $s_{ij}(\mathbf{d}) = \sum_k d_{ik}d_{kj}$. As shown in equation 3.5, individuals are only going to form connections if their idiosyncratic shock is low enough. In particular, following definition 6, we can divide the space of each u_{ij} shock into three components given by two outer buckets and one inner bucket. Because there is only one possible intransitive triad for each dyad, then $\bar{s} = 1$. Figure 3.1 shows the set of all possible buckets for individuals 1, 2, and 3, where for simplicity, we consider a bounded space of $\mathbf{u}$. Panel (a) shows that the different cubes represent 27 possible buckets in the three-dimensional graph, associated with either unique or multiple equilibria (depending on whether the cubes are in the outer or inner buckets). Panel (b) in the same figure shows all the possible equilibrium networks (up to isomorphisms).*

Figure 3.1: Multiplicity Regions and Possible Equilibrium Networks



Panel (a) Panel (b)

Note: This example shows that multiplicity regions are composed by buckets. Moreover, it exemplifies how multiple equilibria can arise when the utility function (3.1) includes externalities of the form $s_{ij}(\mathbf{d}) = \sum_k d_{ik}d_{kj}$. Panel (a) displays the 3D space of the errors u_{ij} for individuals $1, 2$, and 3. The blue cubes represent different buckets that form a multiplicity region where only the empty network is an equilibrium. Panel (b) shows all the possible equilibrium networks -up to isomorphisms- for this simplified example with three individuals.

To exemplify the relationship between buckets and multiplicity regions, Panel (a) highlights in blue the cubes representing outer buckets for all individuals in which only the empty network is an equilibrium outcome. Those blue buckets are all part of the same multiplicity region (where only the empty network is an equilibrium outcome). In addition to being generated by the outer buckets, the empty network can also arise from an inner bucket with the potential to generate multiple equilibria. For instance, the empty network can also occur in the most inner cube (composed of inner buckets), where the empty and complete networks are possible equilibria. This example shows that the same equilibrium outcome can belong to different multiplicity regions composed of different buckets.

The fact that multiplicity regions are made up of buckets guarantees invariability of the shape of the areas for small changes in the parameters θ. Intuitively, changes in the parameters θ can only affect the shape of the buckets, not the number of partitions of $\mathbb{R}$. Therefore, the number of multiplicity regions does not change when the parameters change. The above discussion matters for identification because the invariability of the number of multiplicity regions to changes in the parameters θ implies that the selection probabilities are independent of changes in the parameters. The following Proposition shows that under the proposed framework, the likelihood in (3.7) is locally identified.

Theorem 5 (Identification)**.** *Let Assumption 18 hold. Then, the true vector of payoff parameters θ_0 and the selection probabilities $h_{D0}(\mathbf{d})$ are locally identified for all $\mathbf{d}$ from the likelihood $P(\mathbf{d};\theta,h_0)$ at (θ_0, f_{D0}) if the following matrix is full rank*

$$\sum_m \begin{bmatrix} \frac{\partial G(\mathbf{d}_{N,1},m;\theta)/\partial\gamma}{\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell})} & -\frac{G(\mathbf{d}_{N,1},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} & \cdots & -\frac{G(\mathbf{d}_{N,1},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} & \cdots & -\frac{G(\mathbf{d}_{N,1},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} \\ \vdots & \vdots & \vdots & \vdots & & \\ \frac{\bar{h}_D(\mathbf{d}_{N,j})\partial G(\mathbf{d}_{N,j},m;\theta)/\partial\gamma}{\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell})} & -\frac{\bar{h}_D(\mathbf{d}_{N,j})G(\mathbf{d}_{N,j},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} & \cdots & \frac{(\sum_{i\neq j}\bar{h}_D(\mathbf{d}_{N,i}))G(\mathbf{d}_{N,j},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} & \cdots & -\frac{\bar{h}_D(\mathbf{d}_{N,j})G(\mathbf{d}_{N,j},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} \\ \vdots & \vdots & \vdots & \vdots & & \\ \frac{\bar{h}_D(\mathbf{d}_{N,J})\partial G(\mathbf{d}_{N,J},m;\theta)/\partial\gamma}{\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell})} & -\frac{\bar{h}_D(\mathbf{d}_{N,J})G(\mathbf{d}_{N,J},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} & \cdots & -\frac{\bar{h}_D(\mathbf{d}_{N,J})G(\mathbf{d}_{N,J},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} & \cdots & \frac{(\sum_{i\neq J}\bar{h}_D(\mathbf{d}_{N,i}))G(\mathbf{d}_{N,J},m;\theta)}{(\sum_{\ell\in m}\bar{h}_D(\mathbf{d}_{N,\ell}))^2} \end{bmatrix}$$

where $J=|\mathbb{D}|$, and we define $\bar{h}_D(\mathbf{d}_i)=h_D(\mathbf{d}_{N,i})/h_D(\mathbf{d}_{N,1})$ as the relative probability of $\mathbf{d}_{N,i}$ with respect to $\mathbf{d}_{N,1}$. We can only identify the relative selection probabilities because we have the additional condition that $\sum_{\mathbf{d}\in\mathbb{D}} h_0(\mathbf{d}) = 1$. We chose the normalizing probability to be $h_D(\mathbf{d}_{N,1})$ with out lost of generality.

Theorem 5 shows that, in general, for an arbitrarily large population, the likelihood in 3.7 contains enough information to point-identify both the parameters of interest and selection probabilities. However, in practice, it is unfeasible to construct a probability distribution over a sequence of relatively large networks. The issue is that the space of networks configurations is excessively large even for a moderate number of nodes, such that finding two networks that are isomorphic becomes impractical.[3] To circumvent the issue, we propose to modify the likelihood function in 3.7 to form a more tractable object. Details are given in Section 3.3.2.

3.3 Bayesian Algorithm

Before proceeding with a complete description of the Bayesian algorithm, it is informative to understand the simple case with $\gamma_0 = 0$. This will allow us to set the stage and introduce our assumptions to obtain a tractable Bayesian specification of the estimation problem. In this description, we work under the more general model where there are two sources of unobserved heterogeneity $\mathbf{A}$ and $\mathbf{B}$, but the procedure also applies to the special case where $\mathbf{A} = \mathbf{B}$. These assumptions concern the unobservable A and B as well as their relationships to covariates X. Specifically, in this context Pelican and Graham (2020) essentially treat A and B as fixed effects in the sense that no assumptions are made about the distribution of (A, B) conditional on X. On the other hand, we take a correlated random effects approach (Mundlak, 1978; Chamberlain, 1982) that maintains the essential features of the problem while making estimation more straightforward, particularly in a Bayesian framework.

[3] Obtaining an empirical analogue of a probability distribution over networks would require observing the same network configurations many times to compute relative frequencies, which becomes even more impractical.

3.3.1 Simplified Version

Our results so far do not depend on the specific form of the distribution of idiosyncratic errors $f_U(\mathbf{u})$. As mentioned previously, while a standard assumption in the network formation literature is to specify this distribution as logistic, we instead assume that u is i.i.d. across dyads with a standard normal distribution such that

$$f_U(u) = \prod_{i \neq j} \phi(u_{ij}) \, ,$$

where $\phi(\cdot)$ is the standard normal density. Setting the variance of this distribution to unity can be seen as an identifying restriction, which is standard in models with binary dependent variables (see, e.g., pp. 476, Cameron and Trivedi, 2005).[4] Indeed, as we only observe whether a specific link was formed or not, all scalings of the idiosyncratic shocks will be observationally equivalent.

Under our equilibrium concept, when $\gamma_0 = 0$ the equilibrium selection distribution is degenerate and places all of its mass at the only equilibrium network d that satisfies

$$d_{ij} = d_{ji} = \mathbb{I}(A_i + B_j + W_{ij}'\lambda \geq u_{ij})\mathbb{I}(A_j + B_i + W_{ji}'\lambda \geq u_{ji}) \; i = 1, \ldots, N; \; j = 1, \ldots, N; \; i < j.$$

We introduce the following two assumptions on the distribution of (A, B) conditional on X.

Assumption 19 (Conditionally independent)**.** *For all $i = 1, \ldots, N$ and $j = 1, \ldots, N$ with*

[4] The use of a standard logistic density in the network formation literature, which sets the variance of u_{ij} equal to $\pi^2/3$, is another such identifying restriction. Of course, other restrictions are possible, such as fixing the value of one coefficient or the sum of the coefficients. However, these restrictions would impact the meaning and interpretation of preference parameters, and so we choose to fix the variance of shocks instead.

$i \neq j$, it holds that (i) $(A_i, B_i) \perp (A_j, B_j) | X_i, X_j$ and (ii) $f(A_i, B_i | X_i, X_j) = f(A_i, B_i | X_i)$.

Assumption 20 (Correlated random effects)**.** *For all i, we have $(A_i, B_i)|X_i \sim \mathcal{N}(\Phi X_i, \Sigma)$, where $\Phi = [\phi_1 \, \phi_2]'$ is a $2 \times K$ matrix of coefficients and Σ is a 2×2 covariance matrix.*

Assumption 19.(i) simply states that the in- and out-degree heterogeneity effects are independent across individuals once you condition on the covariates of each dyad. Assumption 19.(ii) then states that the joint distribution of these effects will only depend on each individual's covariates. Finally, Assumption 20 is a correlated random effects specification for a network framework, similar to that in Mundlak (1978). This device is widely used in nonlinear panel data methods in order to deal with unobserved heterogeneity (see, e.g., Chapter 11 of Wooldridge, 2010). In fact, Assumption 19.(i) together with linearity in the conditional expectation of A_i and B_i, which is part of Assumption 20, is enough to satisfy Assumption 19.(ii) without loss of generality. To see this, note that a more general device such as that in Chamberlain (1980) would set $(A_i, B_i)|X_i, X_j \sim \mathcal{N}(\Phi_1 X_i + \Phi_2 X_j, \Sigma)$ for each dyad, where $\Phi_1 = [\phi_{11} \, \phi_{12}]'$ and $\Phi_2 = [\phi_{21} \, \phi_{22}]'$. However, as A_i and B_i never appear together in the marginal utility equation for any individual i, we cannot identify both Φ_1 and Φ_2 separately; we can only identify their sum. Thus, we do not gain more flexibility by including both sets of covariates.

Define $u^*_{ij} \equiv A_i + B_j + W'_{ij}\lambda - u_{ij}$ for $i = 1, \ldots, N$ and $j = 1, \ldots, N$ with $i \neq j$. Note that, using Assumption 20, we can write $A_i = X'_i\phi_1 + a_i$ and $B_i = X'_i\phi_2 + b_i$ for all i, where $(a_i, b_i) \sim \mathcal{N}(0, \Sigma)$ and (a_i, b_i) is independent of X. Replacing into the definition of u^*_{ij}, we have

$$u^*_{ij} = X'_i\phi_1 + X'_j\phi_2 + W'_{ij}\lambda + a_i + b_j - u_{ij} \, . \tag{3.8}$$

Following the way the elements of an adjacency matrix $d \in \mathbb{D}^N$ are indexed, we can first stack (3.8) across i for a given j to obtain

$$u^*_{-j} = X_{-j}\phi_1 + \iota_{N-1}X_j'\phi_2 + W_{-j}\lambda + a_{-j} + \iota_{N-1}b_j - u_{-j}, \quad j = 1, \ldots, N;$$

where ι_{N-1} is an $(N-1)$-dimensional vector of ones, u^*_{-j} and a_{-j} are $N-1$-dimensional vectors, X_{-j} is a $(N-1) \times K$ matrix, W_{-j} is a $(N-1) \times K^2$ matrix and we define

$$u^*_{-j} = \begin{bmatrix} u^*_{1j} \\ \vdots \\ u^*_{j-1,j} \\ u^*_{j+1,j} \\ \vdots \\ u^*_{Nj} \end{bmatrix}, X_{-j} = \begin{bmatrix} X_1' \\ \vdots \\ X_{j-1}' \\ X_{j+1}' \\ \vdots \\ X_N' \end{bmatrix}, W_{-j} = \begin{bmatrix} W_{1j}' \\ \vdots \\ W_{j-1,j}' \\ W_{j+1,j}' \\ \vdots \\ W_{Nj}' \end{bmatrix}, a_{-j} = \begin{bmatrix} a_1 \\ \vdots \\ a_{j-1} \\ a_{j+1} \\ \vdots \\ a_N \end{bmatrix}, u_{-j} = \begin{bmatrix} u_{1j} \\ \vdots \\ u_{j-1,j} \\ u_{j+1,j} \\ \vdots \\ u_{Nj} \end{bmatrix}.$$

We can then stack across the index j to obtain

$$u^* = FX\phi_1 + GX\phi_2 + W\lambda + Fa + Gb - u\,,$$

where u^* and u are $N(N-1)$-dimensional vectors, a and b are N-dimensional vectors, F and G are $N(N-1) \times N$ matrices, X is a $N \times K$ matrix, W is a $N(N-1) \times K^2$ matrix,

$$u^* = \begin{bmatrix} u^*_{-1} \\ \vdots \\ u^*_{-N} \end{bmatrix}, F = \begin{bmatrix} F_{-1} \\ \vdots \\ F_{-N} \end{bmatrix}, X = \begin{bmatrix} X'_1 \\ \vdots \\ X'_N \end{bmatrix}, W = \begin{bmatrix} W_{-1} \\ \vdots \\ W_{-N} \end{bmatrix}, a = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix},$$

$G = I_N \otimes \iota_{N-1}$ and F_{-j} is an identity matrix of order N with column j removed. Using basis vectors $e_i \in \mathbb{R}^N$ that have a 1 at component i and zeros everywhere else, we can write $F_{-j} = [e_1 \cdots e_{j-1}\, e_{j+1} \cdots e_N]'$.

Finally, we can condense the resulting expression by defining $\widetilde{X} = [F X\, GX\, W]$, $H = [F\, G]$, $\beta = [\phi_1'\, \phi_2'\, \lambda']'$ and $c = [a'\, b']'$ to obtain

$$u^* = \widetilde{X}\beta + Hc - u\,. \tag{3.9}$$

With these definitions, we see that the unique equilibrium network d satisfies $d_{ij} = d_{ji} = \mathbb{I}(u^*_{ij} \geq 0) \cdot \mathbb{I}(u^*_{ji} \geq 0)$. As u^* is a multivariate normal distribution conditional on X, β, c and Σ, this is similar to a multivariate probit structure. Existing Bayesian algorithms such as the data augmentation approach (Albert and Chib, 1993) are designed to deal with such structures. We will take advantage of these methods for the algorithms we propose in this paper. As a first step, we introduce u^* as additional latent variables, which lets us define the likelihood for this simplified problem as

$$P(d|u^*, \beta, c, \Sigma) = \prod_{i=1}^{N} \prod_{j>i} [\mathbb{I}(u^*_{ij} \geq 0) \cdot \mathbb{I}(u^*_{ji} \geq 0)]^{d_{ij}} [1 - \mathbb{I}(u^*_{ij} \geq 0) \cdot \mathbb{I}(u^*_{ji} \geq 0)]^{1-d_{ij}} \tag{3.10}$$

To complete a Bayesian specification of the problem, we let the joint prior distribution be $\pi(u^*, \beta, c, \Sigma) = \pi(u^*|\beta, c, \Sigma)\pi(\beta)\pi(c|\Sigma)\pi(\Sigma)$. First, note that $u^*|\beta, c, \Sigma \sim \mathcal{N}(\widetilde{X}\beta - Hc, I_{N(N-1)})$. Additionally, given Assumption 20, we have $c|\Sigma \sim \mathcal{N}(0, \Sigma \otimes I_N)$. For the remaining components, we will assume the standard conditionally conjugate priors given as

$$\beta \sim \mathcal{N}(\underline{\beta}, \underline{B}),$$

$$\Sigma \sim \mathcal{IW}(\underline{v}, \underline{\Sigma}),$$

where $\mathcal{IW}$ is the inverse Wishart distribution. Symbols with an underline are prior hyperparameters and the updated (posterior) quantities will be denoted using an overline. Combining the likelihood in (3.10) with these priors allows us to find the posterior distribution of all quantities of interest. For the coefficients β and random effects c, we can obtain a joint posterior as $\pi(\beta, c|u^*, \Sigma, d) = \pi(\beta|u^*, \Sigma, d)\pi(c|u^*, \beta, \Sigma, d)$. Standard updates (see, e.g., Chib, 2008) result in posteriors

$$\beta|u^*, \Sigma, d \sim \mathcal{N}(\bar{\beta}, \bar{B}),$$

$$c|u^*, \beta, \Sigma, d \sim \mathcal{N}(\bar{c}, \bar{V}),$$

where

$$\bar{B} = (\widetilde{X}'\Omega^{-1}\widetilde{X} + \underline{B}^{-1})^{-1}\,,$$

$$\bar{\beta} = \bar{B}(\widetilde{X}'\Omega^{-1}u^* + \underline{B}^{-1}\underline{\beta})\,,$$

$$\bar{V} = (H'H + \Sigma^{-1} \otimes I_N)^{-1}\,,$$

$$\bar{c} = \bar{V}H'(u^* - \widetilde{X}\beta)\,,$$

and we define $\Omega = I_{N(N-1)} + H(\Sigma \otimes I_N)H'$. For the posterior distribution of Σ, we introduce a matrix version of c that simplifies the resulting expression; i.e., we define $C = [a\, b]$ such that $\text{vec}(C) = c$, where vec is the vectorization operator. Using this definition, we can obtain $\Sigma|u^*, \beta, c, d \sim \mathcal{IW}(\bar{v}, \bar{\Sigma})$, where $\bar{v} = \underline{v} + N$ and $\bar{\Sigma} = C'C + \underline{\Sigma}$.

Finally, we need the posterior of u^* conditional on the parameters and network d. From (3.10), and given the conditional independence of u^* across dyads, we have that for all $i = 1, \ldots, N$ and $j = 1, \ldots, N$ with $i < j$

$$(u^*_{ij}, u^*_{ji})|\beta, c, \Sigma, d \sim \begin{cases} \mathcal{TN}_{[0,\infty)\times[0,\infty)}(\mu_{ij}, I_2) & \text{if } d_{ij} = d_{ji} = 1\,, \\ \mathcal{TN}_{[\mathbb{R}\times(-\infty,0)]\cup[(-\infty,0)\times(0,\infty)]}(\mu_{ij}, I_2) & \text{if } d_{ij} = d_{ji} = 0\,, \end{cases}$$

where $\mathcal{TN}_{\mathcal{S}}$ is a bivariate normal distribution truncated to set S, we define $\mu_{ij} = [\widetilde{x}'_{ij}\beta - a_i - b_j, \widetilde{x}'_{ji}\beta - a_j - b_i]$ and $\widetilde{x}_{ij}$ is the ij-th row of $\widetilde{X}$. As we have all conditional posterior distributions for the quantities of interest, we can set up a Gibbs sampling algorithm to obtain draws from the joint posterior distribution of interest.

3.3.2 Full algorithm

The previous section assumed there were no externality effects by setting $\gamma_0 = 0$ and obtained the full posterior distributions of the preference parameters. However, this paper aims to conduct inference in cases where $\gamma_0 \geq 0$, which, as mentioned previously, generates externalities and complicates the estimation process. Like the previous section, we rely on data augmentation by including the latent u^* as part of our sampling scheme. Note that, using (3.8), we can express the marginal utilities in (3.2) as

$$\Delta v_{ij}(\mathbf{d}_{-ij}; u^*_{ij}) = u^*_{ij} + \gamma s_{ij}(\mathbf{d}) \tag{3.11}$$

In this way, we can separate the effects of degree-heterogeneity and homophily on marginal utility from those of the externalities. Furthermore, we see that this simplifies our definitions for buckets and multiplicity regions given in Section 3.2. That is, we can simply re-define buckets and multiplicity regions from the support of original shocks $\mathbf{u}$ to the transformed $\mathbf{u}^*$. This change in definition modifies the buckets to be of the form $(-\infty, \gamma\underline{s}]$, $(\gamma\bar{s}, \infty)$, or $(\gamma s_k, \gamma s_{k+1}]$ for $\underline{s} \leq s_k < \bar{s}$, where μ_{ij} is no longer relevant (this change in expectation is included as part of the distribution of $\mathbf{u}^*$ after the transformation). We can also update our definitions of the equilibrium selection distribution $\mathcal{N}(\cdot)$, as, conditional on $\mathbf{u}^*$, all of the multiplicity in equilibria come from the dyad-level shocks u^*_{ij} falling into regions that only depend on γ and the externality functions $s_{ij}(\mathbf{d})$. With this change, this equilibrium selection distribution becomes

$$\mathcal{N}(\mathbf{d}, \mathbf{u}; \theta) = \mathcal{N}(\mathbf{d}, \mathbf{u}^*; \gamma) = \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}_m} h_0(\mathbf{d}')} \mathbb{I}(\mathbf{u}^* \in m) g(\mathbf{d}, \mathbf{u}^*; \gamma) \tag{3.12}$$

where now

$$\begin{aligned} g(\mathbf{d}, \mathbf{u}^*; \gamma) = \prod_{i=1}^{N} \prod_{j>i} & \left[\mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) \geq -u^*_{ij}\right) \mathbb{I}\left(\gamma s_{ji}(\mathbf{d}) \geq -u^*_{ji}\right)\right]^{d_{ij}} \\ & \times \left[1 - \mathbb{I}\left(\gamma s_{ij}(\mathbf{d}) \geq -u^*_{ij}\right) \mathbb{I}\left(\gamma s_{ji}(\mathbf{d}) \geq -u^*_{ji}\right)\right]^{1-d_{ij}} \end{aligned} \tag{3.13}$$

These simple transformations arise because the number of multiplicity regions and their relative ordering is independent of the preference parameters other than γ.[5] Using these definitions, we can re-express the joint distribution of $\mathbf{d}$ and $\mathbf{u}^*$ as

$$P(\mathbf{d}, \mathbf{u}^*; \theta) = \mathcal{N}(\mathbf{d}, \mathbf{u}^*; \gamma) f(\mathbf{u}^* | \beta, \mathbf{c}, \Sigma) \tag{3.14}$$

where the conditional distribution $f(\mathbf{u}^*|\beta, \mathbf{c}, \Sigma)$ is multivariate normal as found in the simplified version of the algorithm. Given the addition of γ, we modify our priors accordingly. That is, we now specify $\pi(\mathbf{u}^*, \beta, \mathbf{c}, \Sigma, \gamma) = f(\mathbf{u}^*|\beta, \mathbf{c}, \Sigma)\pi(\beta)\pi(\mathbf{c}|\Sigma)\pi(\Sigma)\pi(\gamma)$, where $\pi(\gamma)$ is the prior distribution for the externality parameter γ. As we assume in our identification results that $\gamma \geq 0$, we can incorporate this restriction into our prior $\pi(\gamma)$ by choosing a distribution with support on the non-negative reals (e.g., a gamma distribution).

A crucial consequence of including the latent $\mathbf{u}^*$ into our sampling scheme is that now the likelihood (3.14) separates into two terms. The first term depends on the parameters only

[5] Indeed, $\mathbf{c}$ and β control the translation of each bucket through μ_{ij}, while γ controls their widths.

through γ, while the other preference parameters enter the likelihood through the second term. This immediately implies that the conditional posterior distributions of $\beta, \mathbf{c}$ and Σ will remain the same as in the simplified version of the algorithm given in the previous subsection. To see this, note that the joint conditional posterior is

$$\begin{aligned}\pi(\beta, \mathbf{c}, \Sigma | \mathbf{d}, \mathbf{u}^*, \gamma) &\propto \mathcal{N}(\mathbf{d}, \mathbf{u}^*; \gamma) f(\mathbf{u}^* | \beta, \mathbf{c}, \Sigma) \pi(\beta) \pi(\mathbf{c}|\Sigma) \pi(\Sigma) \pi(\gamma) \\ &\propto f(\mathbf{u}^* | \beta, \mathbf{c}, \Sigma) \pi(\beta) \pi(\mathbf{c}|\Sigma) \pi(\Sigma)\end{aligned}$$

Thus, Gibbs sampling for $\beta, \mathbf{c}$ and Σ proceeds exactly as before once we obtain values of $\mathbf{u}^*$ and γ. To complete our sampling scheme, we need to obtain the conditional posteriors for $\mathbf{u}^*$ and γ. For the former, write

$$\pi(\mathbf{u}^* | \mathbf{d}, \beta, \mathbf{c}, \Sigma, \gamma) \propto \mathcal{N}(\mathbf{d}, \mathbf{u}^*; \gamma) f(\mathbf{u}^* | \beta, \mathbf{c}, \Sigma) \tag{3.15}$$

$$= \sum_{m \in M} \frac{h_0(\mathbf{d})}{\sum_{\mathbf{d}' \in \mathbb{D}_m} h_0(\mathbf{d}')} \mathbb{I}(\mathbf{u}^* \in m) g(\mathbf{d}, \mathbf{u}^*; \gamma) f(\mathbf{u}^* | \beta, \mathbf{c}, \Sigma) \tag{3.16}$$

To make sense of this expression, first note that from (3.13), $g(\mathbf{d}, \mathbf{u}^*; \gamma)$ imposes constraints on each pair (u^*_{ij}, u^*_{ji}), as they need to be such that they satisfy our equilibrium concept (for a given equilibrium network $\mathbf{d}$). Furthermore, this latent component of marginal utility should be such that equilibrium solutions arising from $\mathbf{u}^*$ (and conditional on γ) belong to the same multiplicity region as $\mathbf{d}$. Finally, as $\mathbf{d}$ can be an equilibrium network for several multiplicity regions, the posterior of $\mathbf{u}^*$ weights each of these regions according to the underlying network distribution from Assumption 18. Putting everything together implies that the posterior of $\mathbf{u}^*$ is a mixture of truncated multivariate normals across multiplicity regions, with mixture weights given by the conditional probability of observing network $\mathbf{d}$ in each multiplicity region.[6]

For the posterior of γ, we can similarly obtain

$$\pi(\gamma|\mathbf{d}, \mathbf{u}^*, \beta, \mathbf{c}, \Sigma) = \pi(\gamma|\mathbf{d}, \mathbf{u}^*) \propto \mathcal{N}(\mathbf{d}, \mathbf{u}^*; \gamma)\pi(\gamma) \tag{3.17}$$

This equation in a way evidences the inherent identification problem for γ, as conditional on $\mathbf{u}^*$, all identifying data information comes from the associated probabilities for each multiple equilibria in $\mathcal{N}(\mathbf{d}, \mathbf{u}^*; \gamma)$. The remaining information comes purely from the prior.

The main issue in obtaining draws for $\mathbf{u}^*$ and γ is that we must now find a way to evaluate these equilibrium selection probabilities; i.e., we need to compute $\mathcal{N}(\mathbf{d}, \mathbf{u}^*; \gamma)$. As this is unfeasible given the current setup, we introduce one further simplifying assumption in the form of a composite likelihood that replaces the full likelihood (3.14) for sampling the two remaining components.

3.3.3 Composite Likelihood Function

Let $\mathbf{d}_{K,k}$ be the subgraph of size K induced by the network $\mathbf{d}$. Construct the sequence of N choose K subnetworks $\{\mathbf{d}_{K,1}, \dots, \mathbf{d}_{K,N_K}\}$, where N_K is the total number of subnetworks that can be formed by forming all possible combinations of nodes of size K. Therefore, the marginal probability of observing network $\mathbf{d}$ can be constructed by the joint probability of observing the sequence of subnetworks $\{\mathbf{d}_{K,1}, \dots, \mathbf{d}_{K,N_K}\}$. Under the assumption that each subnetwork $\mathbf{d}_{K,k}$ forms following the Pairwise Stability criteria in 4, the marginal likelihood

[6] If a network $\mathbf{d}$ cannot arise as an equilibrium network for shocks in multiplicity region m (i.e., $\mathbf{d} \notin \mathbb{D}_m$), then the associated mixture weight of multiplicity region m is 0 because $\mathbf{u}^* \notin m$.

of each subgraph will be given by

$$P(\mathbf{d}_{K,k}, \mathbf{u}_{K,k}; \theta, h_0) = \sum_{m \in M} \frac{h_0(\mathbf{d}_{K,k})}{\sum_{\mathbf{d}'_{K,k} \in \mathbb{D}_m} h_0(\mathbf{d}'_{K,k})} \mathbb{I}(\mathbf{u}^*_{K,k} \in m) g(\mathbf{d}_{K,k}, \mathbf{u}^*_{K,k}; \gamma),$$

for all the possible multiplicity regions M for subnetworks of size K, and for all k in the sequence of N_K subnetwokrs. Characterizing the joint distribution over the sequence $\{\mathbf{d}_{K,1}, \ldots, \mathbf{d}_{K,N_K}\}$ is also unfeasible. Instead, we follow the idea in Graham (2020) and propose a composite likelihood that is formed by multiplying the marginal distributions of N_K subgraphs forming $\mathbf{d}N$

$$P(\mathbf{d}, \mathbf{u}^*; \theta, h_0) = \prod_{k=1}^{N_K} P(\mathbf{d}_{K,k}, \mathbf{u}^*_{K,k}; \theta, h_0). \tag{3.18}$$

The advantage of a composite likelihood is that, even though it fails to represent the dependence structures across different subgraphs correctly, if the marginal distributions are correctly specified, estimators for θ based on the composite likelihood are consistent for the true population parameter (Cox and Reid, 2004; Varin et al., 2011). In addition, by choosing values of K that make the space of potential network configurations manageable, it is possible to provide a tractable probability distribution over the set of possible equilibrium networks. Interestingly, when $K \to N$, the composite likelihood in (3.18) collapses to (3.7), and we are back to the situation where the identification of selection probabilities is an intractable problem.

3.4 Empirical Application

This section presents an empirical application that highlights the relevance and practicality of our proposed Bayesian estimator. We use data from Banerjee et al. (2013), whose primary objective is to study learning diffusion through social networks in the context of microfinance participation for 77 villages in Karnataka, India. Banerjee et al. (2013) examines microfinance take-up diffusion in a network, arguing that the information seed is exogenous because it is decided by a third-party institution that is not explicitly maximizing any individual or aggregate objective function. Interestingly, though, the authors acknowledge that the social network is likely endogenous. They mention homophily reasons arguing that connected individuals tend to exhibit strong similarities. Additional strategic considerations can affect individuals' decisions to form connections, such as degree heterogeneity and payoff externalities. Given the availability of granular and rich network information, our data is particularly well-suited to investigate which channels are relevant for individuals to form social ties.

3.4.1 Network Data

The data comes from the Social Networks and Microfinance project, which contains the publicly available data of participation in a program of Bharatha Swamukti Samsthe (BSS), a microfinance institution (MFI) in rural southern Karnataka. The data were collected in 2006 for 77 villages and included information of thirteen possible dimensions in which individuals can be connected, including visiting each other, going to pray, borrowing and lending money and goods, obtaining advice, and giving advice (Banerjee et al., 2013). The data contain a village questionnaire and a complete census, including information on all villages' households. Individuals' characteristics include gender, age, religion, caste, subcaste, mother tongue, whether the individual is a village native, education, work frequency, and occupations. There are two levels of aggregation in the data: individuals and households. Given that our objective is to estimate the utility function parameters of individuals making strategic choices, we will use the individual-level data to fit our model.

Regarding the construction of the social network, we follow Banerjee et al. (2013) when defining our network of interest. We consider the connections to be undirected, so we use the symmetric version of the adjacency matrices capturing the relationships between individuals. Therefore, two individuals are neighbors in the network if at least one mentions the other as a contact in response to some network question. Finally, instead of considering each of the thirteen possibles ways people interact in these villages in a different network, we consider two people linked if they mentioned each other in at least one type of relationship. Table 3.2 presents some summary statistics for the social networks of the 20 most populated villages in the sample. The networks consistently show relatively high average degrees between approximately 8 and 10 connections. The levels of transitivity are across the board lower than 0.5, meaning that there are more incomplete triangles than complete ones. These transitivity values present initial evidence that individuals may not have a substantial payoff from completing triangles (Jackson and Rogers, 2007). These networks also feature a relatively high average shortest path length (distance) and many disconnected components. All these attributes suggest highly clustered networks that are sparse across clusters.

Table 3.2: Summary Network Statistics

Village	Nodes	Average Degree	Transitivity	Average Distance	Components
60	1775	8.67	0.45	4.63	16
28	1612	9.40	0.52	4.36	18
59	1599	8.37	0.44	4.68	20
52	1525	10.28	0.36	4.13	12
71	1387	8.26	0.33	4.44	13
3	1380	7.78	0.35	4.36	21
39	1339	8.96	0.45	4.29	14
29	1337	7.61	0.39	4.53	21
65	1331	9.22	0.30	4.09	11
25	1313	9.33	0.46	4.21	10
64	1286	8.67	0.46	4.46	10
46	1257	7.71	0.39	4.55	12
23	1252	8.40	0.39	4.27	19
36	1214	8.61	0.31	4.07	21
32	1181	9.42	0.42	4.10	15
55	1180	7.84	0.48	4.84	12
76	1154	8.15	0.45	4.48	13
18	1146	9.10	0.38	4.13	5
19	1134	9.19	0.40	4.18	5
40	1097	7.89	0.40	4.58	11

Note: table with the network statistics for the 20 most populated villages out of the 77 villages in the data. The average distance is calculated as the maximum distance in all possible connected components.

3.4.2 Data on Individual Characteristics

The Social Networks and Microfinance project also includes a battery of variables characterizing individuals in the villages. We have information on gender, individuals' role in the household (head of the household, spouse of the head, or other), age, religion, cast, sub-cast, languages that the individuals speak, working status, saving behavior, and participation in the financial market. We choose a subset of those characteristics to construct our homophily measure in equation $W_{i,j}$ from Equation (3.5). In particular, we use the working status, gender, individuals' role in the household, cast, and whether or not the individual is native from the village and construct a set of dummy variables. Using this information, the measure of homophily is whether two individuals match in the value of each of the dummy variables. If they match, the homophily variable takes the value of zero and the value of one if they mismatch.

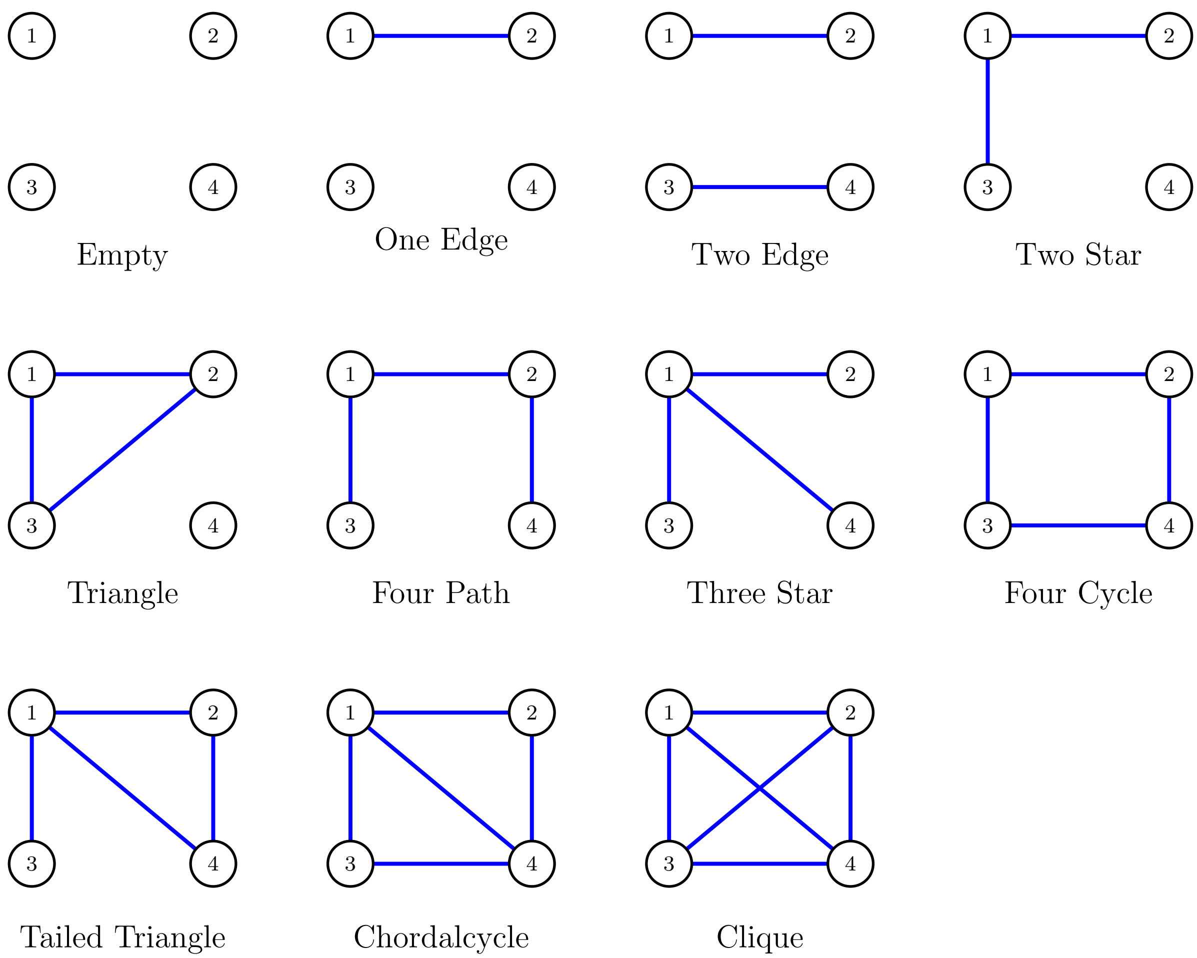
3.4.3 Subgraphs and Selection Probabilities

As we argue in Section 3.2.3, it is untractable to estimate probability distribution over a sequence of large networks in practice. The data we are using contains large networks for each village, with a sample size between 780 and 1775 nodes. Section 3.3.3 presents a solution for this practical issue based on the idea of subgraphs and composite likelihoods. One relevant parameter that we need to choose is the number of subgraphs K. As argued by Graham (2017), working with subgraphs of size four (tetrads) provides enough variation to identify homophily parameters while at the same time having the the advantage of yielding a criterion function that is easy to evaluate and maximize. Tetrads subgraphs also contain enough nodes to provide variation in the number of complete and incomplete triads. Based on these arguments, we choose $K = 4$ to perform our Bayesian estimation algorithm. An additional convenient feature of working with tetrads, is that it is possible to completely characterize all the possible configuration of the subnetworks. Figure 3.2 presents the 11 unique isomorphism classes in which tetrads can be wired.

Selection Probability Estimator

Based on the 11 isomorphism classes in Figure 3.2 and under Assumption 18, it is possible to non-parametrically estimate the selection probabilities for each isomorphism given the observation of one large network. We compute the estimator as follows: given the observations of a sampled network $\mathbf{d}$ of size n, construct the sequence of n choose 4 possible tetrads in the network and calculate the frequency of each isomorphism. The frequency values are our estimator for the predetermined probabilities $h_0(\mathbf{d}_{4,k})$. After estimating the predetermined probabilities, conditional on the form of the externality component $s_{ij}(\mathbf{d})$, it is possible to form the buckets as defined in 6, construct the multiplicity region following proposition 2, and form the quotient $h_0(\mathbf{d}_{4,k})/\sum_{d'\in\mathbb{D}_m}$ for each multiplicity region m. For instance, if $s_{ij}(\mathbf{d})$ is the test for completing triangles externality, for each tetrad, there are a total of three values for the possible number of intransitive triads for individual i. Following the same idea as in equation 3.4, it is possible to construct four buckets. Given that the links are undirected, there are six possible ways in which the tetrads adjacency matrices' upper triangular elements can be arranged that can be mapped to the isomorphism. Therefore, there are 4,096 possible combinations of matrix configurations and buckets, and we can know exactly to what bucket each isomorphism belongs.

Figure 3.2: Tetrads Isomorphisms



Note: each of the graphs represent one possible tetrad configuration. A tetrad can be wired in up to 64 different ways, but the 11 configurations in this figure are the unique isomorphism classes. Any tetrad can be represented by one of the graphs in this figure up to rotating indexes (Graham, 2017).

Table 3.3: Estimators of Tetrads Probabilities

Tetrad	Village 60	Village 28	Village 59
Empty	0.324	0.352	0.358
One Edge	0.449	0.471	0.476
Two Edge	0.049	0.049	0.048
Two Star	0.061	0.047	0.047
Triangle	0.068	0.064	0.056
Four Path	0.009	0.006	0.006
Three Star	0.002	0.001	0.001
Four Cycle	0.000	0.000	0.000
Tailed Triangle	0.007	0.005	0.005
Chordalcycle	0.001	0.001	0.001
Clique	0.029	0.004	0.002

Note: table with the non-parametric estimates of the tetrads probabilities based on the three villages with larger sample sizes. We compute the n choose 4 sequence of tetrads where n represents the number of nodes in each village, and then use that sequence to calculate the proportions of each tetrad isomorphisms.

Table 3.3 presents the estimates of the predetermined probability of each tetrad isomorphism for the three villages with the larger sample sizes. The one Edge is the most likely configuration to emerge, followed by the Empty isomorphisms. These two network configurations are the most sparse among all possible isomorphisms, suggesting that our sample's social networks are relatively sparse. This result is consistent with the relatively low transitivity index and the large average distance from the network statistics presented in Table 3.2. Based on the predetermined probabilities shown in Table 3.3, it is then possible to calculate the relative frequency of each isomorphism with respect to the other network configurations that belong to its multiplicity region. Again, that calculation depends entirely on the specification of the payoff externality form in $s_{ij}(\mathbf{d})$.

3.4.4 Empirical Results

This section presents the empirical results of estimating the network formation rule described in section 3.2.1 without the presence of payoff externalities. These estimators can be seen as the Bayesian-Correlated Effects version of the tetrads fixed effects maximum likelihood estimator proposed by Graham (2017). We focus on the results for the homophily parameters where we fit the network formation model using the observed characteristics described in section 3.4.2. Table 3.4 presents the means and standard deviations of the homophily parameters for work status, gender, head of the household, spouse of the head, casts, and native variables. We focus only on village 60, which contains the largest sample size among all villages.

We find strong evidence of homophily for most of the characteristics in our analysis. The most substantial homophily effects happen among the same gender and working status of villagers. These results have repercussions in the context of learning dynamics because the shape of the network critically depends on the characteristics of the individuals. Therefore, we can expect to see higher levels of information transition across similar people, which can affect policies determining interventions such as injection points (Banerjee et al., 2013).

3.5 Conclusion

This paper provides a novel approach to identify and perform inference on the utility function parameters of a network formation model with payoff externalities. Our identification results rely on the assumption that there is a predetermined probability distribution over the set of all possible equilibrium networks. Under this assumption, we can characterize the selection

Table 3.4: Mean and Standard Deviation for the Posterior Distribution of λ

	Mean	Standard Deviation
Different Work Status	-0.012	0.006
Different Gender	-0.043	0.023
Head of the Household/Others	0.007	0.076
Spouse of the Head/Others	-0.011	0.078
Different Caste	-0.011	0.006
Native and Non-Native	-0.012	0.006

Note: the table presents the mean and standard deviation for the posterior distribution of λ using the Bayesian algorithm described in section 3.3.1. The posterior distribution takes into account the fact that there may exists degree heterogeneity in the network formation rule.

probabilities relying on the fact that the space of idiosyncratic payoff shocks and the type of payoff externalities determine entirely the regions where multiple equilibria happen–we call them *multiplicity regions*. With the characterization of the selection probability, we show that the payoff parameters of interest are identified from the likelihood function that aggregates the likelihoods of observing a network across different multiplicity regions.

We propose a Bayesian algorithm to estimate the parameters of interest. The use of Bayesian methods allows us to sidestep the issue of high-dimensional numerical integration. It also enables us to address the potential problem of statistical dependence in large networks because it does not rely on asymptotic theory to conduct inference. Because we recover the posterior distribution of the parameters of interest, we can perform statistical inference based on that distribution. One potential issue when translating the identification results to the estimation method is that it is unfeasible to construct a probability distribution over a sequence of relatively large networks in practice. We address this issue by building a composite likelihood based on the marginal distribution of all the possible subgraphs forming the network of interest.

Using the proposed estimation methods, we present an empirical application to model the network formation process of individuals creating social connections in villages in Karnataka, India (Banerjee et al., 2013). We characterize the probability distribution of tetrads subgraphs for the social networks and use our Bayesian algorithm to estimate the homophily effects. We find strong evidence of homophily for most of the characteristics in our analysis. The most substantial homophily effects happen among the same gender and working status of villagers.

3.6 Appendix: Proofs of Main Results

Proof of Proposition 2. We prove this proposition by contradiction. Assume there exists a bucket b and two multiplicity regions m_1 and m_2, such that $b \cap m_1 = b_1$ and $b \cap m_2 = b_2$. By the definition of multiplicity region, there must exist some network equilibrium d such that $d \in m_1$ and $d \notin m_2$. Thus, we have $d \in b_1$ and $d \notin b_2$, which is a contradiction. □

Proof of Theorem 5. Let $\bar{h}_D(\boldsymbol{d}_i) = h_D(\boldsymbol{d}_{N,i})/h_D(\boldsymbol{d}_{N,1})$ be the relative probability of $\boldsymbol{d}_{N,i}$ with respect to $\boldsymbol{d}_{N,1}$, where $h_D(\boldsymbol{d}_{N,1})$ was chosen as the normalizing probability without lost of generality. Defining $J = |\mathbb{D}|$, the likelihood function of the problem in equation (3.7) can be written as a system of equation as follows

$$P(\boldsymbol{d}_{1,N}; h_D, \theta) = \sum_{m \in M} \frac{1}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} G(\boldsymbol{d}_{N,1}, m; \theta),$$

$$\vdots$$

$$P(\boldsymbol{d}_{N,j}; h_D, \theta) = \sum_{m \in M} \frac{\bar{h}_D(\boldsymbol{d}_{N,j})}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} G(\boldsymbol{d}_{N,j}, m; \theta),$$

$$\vdots$$

$$P(\boldsymbol{d}_{N,J}; h_D, \theta) = \sum_{m \in M} \frac{\bar{h}_D(\boldsymbol{d}_{N,J})}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} G(\boldsymbol{d}_{N,J}, m; \theta).$$

Without lost of generality, we assume there is one multiplicity region for the rest of the proof. We take the total differentiation of both sides of the system of equations defined before to get

$$
\begin{aligned}
\Delta P(\boldsymbol{d}_{N,1}; h_D, \theta) &= \frac{1}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,i})} \left[\sum_{ij} \frac{\partial G(\boldsymbol{d}_{N,1})}{\partial W_{ij} \lambda} W_{ij} \nabla \lambda + \frac{\partial G(\boldsymbol{d}_{N,1})}{\partial \gamma} \Delta \gamma \right] \\
&\quad - \sum_{l=1}^{J} \frac{G(\boldsymbol{d}_{N,1}; \theta)}{(\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell}))^2} \Delta \bar{h}_D(\boldsymbol{d}_{N,l}) \\
&\vdots \\
\Delta P(\boldsymbol{d}_{N,j}; h_D, \theta) &= \frac{\bar{h}_D(\boldsymbol{d}_{N,j})}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,i})} \left[\sum_{ij} \frac{\partial G(\boldsymbol{d}_{N,j})}{\partial W_{ij} \lambda} W_{ij} \nabla \lambda + \frac{\partial G(\boldsymbol{d}_{N,j})}{\partial \gamma} \Delta \gamma \right] \\
&\quad + \frac{G(\boldsymbol{d}_{N,j}; \theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} \Delta \bar{h}_D(\boldsymbol{d}_{N,j}) \\
&\quad - \sum_{l=1}^{J} \frac{\bar{h}_D(\boldsymbol{d}_{N,j}) G(\boldsymbol{d}_{N,j}; \theta)}{(\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell}))^2} \Delta \bar{h}_D(\boldsymbol{d}_{N,l}) \\
&\vdots \\
\Delta P(\boldsymbol{d}_{N,J}; h_D, \theta) &= \frac{\bar{h}_D(\boldsymbol{d}_{N,J})}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,i})} \left[\sum_{ij} \frac{\partial G(\boldsymbol{d}_{N,J})}{\partial W_{ij} \lambda} W_{ij} \nabla \lambda + \frac{\partial G(\boldsymbol{d}_{N,J})}{\partial \gamma} \Delta \gamma \right] \\
&\quad + \frac{G(\boldsymbol{d}_{N,J}; \theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{i,N})} \Delta \bar{h}_D(\boldsymbol{d}_{N,J}) \\
&\quad - \sum_{l=1}^{J} \frac{\bar{h}_D(\boldsymbol{d}_{N,J}) G(\boldsymbol{d}_{N,J}; \theta)}{(\sum_{\ell \in m} \bar{f}_D(\boldsymbol{d}_{N,i}))^2} \Delta \bar{h}_D(\boldsymbol{d}_{N,l})
\end{aligned}
$$

The identification idea is then to show that when the total derivative of the likelihood function equals zero, the vector of parameters θ and the selection probabilities $h_D(\mathbf{d})$ do not vary. Let $\Delta P(\boldsymbol{d}_j; f_D, \theta) = 0$ for all $j = 1, ..., J$. Defining the following matrix,

$$A = \begin{bmatrix} \frac{\partial G(\boldsymbol{d}_{N,1})}{\partial \gamma} & -\frac{G(\boldsymbol{d}_{N,1};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} & \cdots & -\frac{G(\boldsymbol{d}_{N,1};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,i})} & \cdots & -\frac{G(\boldsymbol{d}_{N,1};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{h}_D(\boldsymbol{d}_{N,j}) \frac{\partial G(\boldsymbol{d}_{N,j})}{\partial \gamma} & -\frac{\bar{h}_D(\boldsymbol{d}_{N,j}) G(\boldsymbol{d}_{N,j};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} & \cdots & \frac{(\sum_{i \neq j} \bar{h}_D(\boldsymbol{d}_{N,i})) G(\boldsymbol{d}_{N,j};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} & \cdots & -\frac{\bar{h}_D(\boldsymbol{d}_{N,j}) G(\boldsymbol{d}_{N,j};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,\ell})} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \bar{h}_D(\boldsymbol{d}_{N,J}) \frac{\partial G(\boldsymbol{d}_{N,J})}{\partial \gamma} & -\frac{\bar{h}_D(\boldsymbol{d}_{N,J}) G(\boldsymbol{d}_{N,J};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,i})} & \cdots & -\frac{\bar{h}_D(\boldsymbol{d}_{N,J}) G(\boldsymbol{d}_{N,J};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,i})} & \cdots & \frac{(\sum_{\ell \neq J} \bar{h}_D(\boldsymbol{d}_{N,i})) G(\boldsymbol{d}_{N,J};\theta)}{\sum_{\ell \in m} \bar{h}_D(\boldsymbol{d}_{N,i})} \end{bmatrix},$$

the system of equations defined before can be represented in matrix form as

$$\mathbf{A} \begin{bmatrix} \Delta\gamma \\ \Delta\bar{h}_D(\boldsymbol{d}_{N,2}) \\ \vdots \\ \Delta\bar{h}_D(\boldsymbol{d}_{N,J}) \end{bmatrix} = \begin{bmatrix} -\sum_{ij} \frac{\partial G(\boldsymbol{d}_{N,1})}{\partial W_{ij}\lambda} W_{ij} \nabla\lambda \\ -\bar{h}_D(\boldsymbol{d}_{N,2}) \sum_{ij} \frac{\partial G(\boldsymbol{d}_{N,2})}{\partial W_{ij}\lambda} W_{ij} \nabla\lambda \\ \vdots \\ -\bar{h}_D(\boldsymbol{d}_{N,J}) \sum_{ij} \frac{\partial G(\boldsymbol{d}_{N,J})}{\partial W_{ij}\lambda} W_{ij} \nabla\lambda \end{bmatrix},$$

Under the full rank assumption the matrix $\mathbf{A}$ is invertible. Given that W_{ij} are random variables, we have that $\Delta P(\boldsymbol{d}_j; f_D, \theta) = 0$ for all $j = 1, ..., J$ if and only if $\nabla\lambda = \mathbf{0}$, and

$$\begin{bmatrix} \Delta\gamma \\ \Delta\bar{h}_D(\boldsymbol{d}_{N,2}) \\ \vdots \\ \Delta\bar{h}_D(\boldsymbol{d}_{N,J}) \end{bmatrix} = \mathbf{0},$$

where $\mathbf{0}$ is a vector of zeros. Therefore, the vector of parameters θ and $h_D(\boldsymbol{d}_N)$ for all $\boldsymbol{d} \in \mathbb{D}_N$ are locally identified from the likelihood $P(\boldsymbol{d}_N; h_D, \theta)$ at (h_{D0}, θ_0). $\square$

3.7 Appendix: Equilibrium in Directed Networks

In this section we show that our identification results are consistent with a network formation model where individuals can choose undirected links. We follow the model by Pelican and Graham (2020) where individuals form directed links based on a payoff function that includes preferences over other individuals' positions in the network. As in the main text, individual's utility from the network configuration is determined by

$$\nu_i\left(\mathbf{d}_i, \mathbf{d}_{-i}; \mathbf{U}\right) = \sum_j d_{ij}\left[A_i + B_j + X_i'\Lambda_0 X_i + \gamma_0 s_{ij}(\mathbf{d}) - U_{ij}\right], \tag{3.19}$$

where the components of the utility are described in the main text (equation (3.1)). We assume a game of complete information where individual i observes $\{A_i, B_i, X_i'\}_{i=1}^N$ and $\{U_{ij}\}_{i\neq j}$, then decides what $N-1$ agents she will send links to. Following the literature of directed network formation, we use the Nash Equilibrium (NE) solution concept. A pure strategy NE corresponds to a strategy $\mathbf{d}^*$ such that $\nu_i\left(\mathbf{d}_i^*, \mathbf{d}_{-i}^*, \mathbf{u}\right) \geq \nu_i\left(\mathbf{d}_i, \mathbf{d}_{-i}^*, \mathbf{u}\right)$.

Based on the marginal utility defined in equation (3.2) in the main text, an observed Nash Equilibrium network is one that satisfies the following $N(N-1)$ set of non-linear equations:

$$d_{ij} = \mathbb{I}\left(A_i + B_j + W_{ij}'\lambda + \gamma_0 s_{ij}(\mathbf{D}) \geq u_{ij}\right),$$

for $i = 1, \ldots, N$ and $i \neq j$. Following Miyauchi (2016) and Pelican and Graham (2020), we define the mapping $\varphi(\mathbf{d}) : \mathbb{D}_N \to \mathbb{I}_{N(N-1)}$ as

$$\varphi(\mathbf{d}) \equiv \begin{bmatrix} \mathbb{I}\left(\Delta\nu_{12}\left(\mathbf{d}_i, \mathbf{d}_{-i}; u_{ij}\right) \geq 0\right) \\ \mathbb{I}\left(\Delta\nu_{13}\left(\mathbf{d}_i, \mathbf{d}_{-i}; u_{ij}\right) \geq 0\right) \\ \vdots \\ \mathbb{I}\left(\Delta\nu_{NN-1}\left(\mathbf{d}_i, \mathbf{d}_{-i}; u_{ij}\right) \geq 0\right) \end{bmatrix},$$

and use Tarski's (1955) fixed point theorem to argue that for $\gamma_0 \geq 0$, and equilibrium exists and the set of all equilibria constitutes a non-empty complete lattice (see Proposition 1 in Miyauchi (2016)). An important insight is that the change in equilibrium concept does not change the idea of multiplicity regions in definition 5. The reason is that multiplicity regions definition takes any arbitrary equilibrium concept as its input, and serves as a way to partition the space of shocks $\mathbf{u}$. Therefore, the likelihood of observing a network $\mathbf{d}_N$ under the Nash Equilibrium concept, can also be written as in equation (3.20) in the main text

$$P(\boldsymbol{d}_N; \theta) = \sum_{m \in M} \frac{h_D(\boldsymbol{d}_N)}{\sum_{i \in m} h_D(\boldsymbol{d}_{N,\ell})} \int_{\boldsymbol{u}_N \in \mathbb{R}^N} \mathbb{I}\{\boldsymbol{u}_N \in m\} g(\boldsymbol{d}_N, \boldsymbol{u}_N; \theta) \prod_{i \neq j} f_U\left(u_{N,ij}\right) \mathrm{d}\boldsymbol{u}_N\,, \quad (3.20)$$

where in this case the function $g(\boldsymbol{d}_N, \boldsymbol{u}_N; \theta)$ takes the form

$$g(\boldsymbol{d}_N, \boldsymbol{u}_N; \theta) = \prod_i \prod_j \mathbb{I}\left(A_i + B_j + W'_{ij}\lambda \geq u_{ij}\right)^{d_{ij}} \times \mathbb{I}\left(A_i + B_j + W'_{ij}\lambda < u_{ij}\right)^{1-d_{ij}}$$

to be consistent with the Nash Equilibrium solution concept. Given that the likelihood function can be still separated into a weighted sum of multiplicity regions probability masses, the identification result in 5 also holds for the directed network formation game.

Bibliography

Albert JH, Chib S. 1993. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* **88**: 669–679.

Alexander C, Piazza M, Mekos D, Valente T. 2020. Peers, schools, and adolescent cigarette smoking. *Journal of Adolescent Health* **29**: 22–30.

Atkisson C, Górski PJ, Jackson MO, Hołyst JA, D'Souza RM. 2020. Why understanding multiplex social network structuring processes will help us better understand the evolution of human behavior. *Evolutionary Anthropology* **29**: 102–107.

Auerbach E. 2022. Identification and estimation of a partially linear regression model using network data. *Econometrica* **90**: 347–365.

Bajari P, Hahn J, Hong H, Ridder G. 2011. A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models. *International Economic Review* **52**: 807–824.

Bajari P, Hong H, Ryan SP. 2010. Identification and estimation of a discrete game of complete information. *Econometrica* **78**: 1529–1568.

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.

Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO. 2013. The diffusion of microfinance. *Science* **341**.

Bayer P, Hjalmarsson R, Pozen D. 2009. Building criminal capital behind bars: Peer effects in juvenile corrections. *The Quarterly Journal of Economics* **124**: 105–147.

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.

Cameron AC, Trivedi PK. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.

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.

Carrell SE, Fullerton RL, West JE. 2009. Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics* **27**: 439–464.

Carrell SE, Hoekstra M, West JE. 2011. Is poor fitness contagious?. evidence from randomly assigned friends. *Journal of Public Economics* **95**: 657–663.

Carrell SE, Sacerdote BI, West JE. 2013. From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica* **81**: 855–882.

Chamberlain G. 1980. Analysis of Covariance with Qualitative Data. *The Review of Economic Studies* **47**: 225–238.

Chamberlain G. 1982. Multivariate Regression Models for Panel Data. *Journal of Econometrics* **18**: 5–46.

Chandrasekhar A. 2016. Econometrics of network formation. In Bramoullé Y, Rogers BW, Galeotti A (eds.) *Oxford Handbook on the Economics of Networks*. Oxford University Press, 45–82.

Chandrasekhar AG, Lewis R. 2016. Econometrics of sampled networks. Unpublished Manuscript.

Chib S. 2008. Panel Data Modeling and Inference: A Bayesian Primer. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, volume 46 of *Advanced Studies in Theoretical and Applied Econometrics*. Springer, 479–515.

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.

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.

Cox DR, Reid N. 2004. A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**: 729–737.

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 A. 2013. Econometric analysis of games with multiple equilibria. *Annu. Rev. Econ.* **5**: 107–131.

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.

Erdös P, Rényi A. 1959. On random graphs. *Publicationes Mathematicae Debrecen* **6**: 290–297.

Estrada J. 2021. Causal inference in multilayer social networks. Unpublished Manuscript.

Estrada J, Huynh KP, Jacho-Chávez DT, Sánchez-Aragón L. 2020. On the identification and estimation of endogenous peer effects in multiplex networks. Unpublished Manuscript.

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

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.

Goyal S, van der Leij MJ, Moraga-González JL. 2006. Economics: An emerging small world. *Journal of Political Economy* **114**: 403–412.

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.

Hajivassiliou VA, McFadden DL. 1998. The method of simulated scores for the estimation of ldv models. *Econometrica* : 863–896.

Humphries M, Gurney K. 2008. Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS One* **3**. ISSN 1932-6203.

Jackson MO. 2010. *Social and economic networks*. Princeton university press.

Jackson MO, Rodriguez-Barraquer T, Tan X. 2012. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review* **102**: 1857–97.

Jackson MO, Rogers BW. 2005. The economics of small worlds. *Journal of the European Economic Association* **3**: 617–627.

Jackson MO, Rogers BW. 2007. Meeting strangers and friends of friends: How random are social networks? *American Economic Review* **97**: 890–915.

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

Jackson MO, Wolinsky A. 1996. A strategic model of social and economic networks. *Journal of Economic Theory* **71**: 44–74. ISSN 0022-0531.

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.

Johnsson I, Moon HR. 2021. Estimation of peer effects in endogenous social networks: Control function approach. *Review of Economics and Statistics* **103**: 328–345.

Joshi S, Mahmud AS, Sarangi S. 2020a. Network formation with multigraphs and strategic complementarities. *Journal of Economic Theory* **188**: 105033.

Joshi S, Mahmud AS, Tzavellas H. 2020b. "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* .

Kreager Da, Rulison K, Moody J. 2020. Delinquency and the structure of adolescent peer groups. *Criminology* **49**: 95–127.

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.

Lewbel A, Qu X, Tang X. 2019. Social networks with misclassified or unobserved links. Unpublished Manuscript.

Liu X, Patacchini E, Zenou Y. 2014. Endogenous peer effects: Local aggregate or local average? *Journal of Economic Behavior and Organization* **103**: 39 – 59.

Manski CF. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* **60**: 531–542.

Manta A, Ho ATY, Huynh KP, Jacho-Chávez DT. 2021. Estimating social effects in a multilayered linear-in-means model with network data. Unpublished Manuscript.

Mas A, Moretti E. 2009. Peers at work. *American Economic Review* **99**: 112–45.

McCulloch RE, Polson NG, Rossi PE. 2000. A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics* **99**: 173–193.

McFadden D. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* : 995–1026.

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.

Milgrom P, Roberts J. 1990. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society* : 1255–1277.

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.

Mundlak Y. 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica* **46**: 69–85.

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.

Pelican A, Graham BS. 2020. An optimal test for strategic interaction in social and economic network formation between heterogeneous agents. Technical report, National Bureau of Economic Research.

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

Sacerdote B. 2001a. Peer effects with random assignment: Results for dartmouth roommates. *Quaterly Journal of Economics* **116**: 681–704.

Sacerdote B. 2001b. Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics* **116**: 681–704.

Sacerdote B. 2014. Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annu. Rev. Econ.* **6**: 253–272.

Salmivalli C. 2020. Bullying and the peer group: A review. *Aggression and Violent Behavior* **15**: 112–120.

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

Tarski A. 1955. A lattice-theoretical fixpoint theorem and its applications. *Pacific journal of Mathematics* **5**: 285–309.

Topkis DM. 1979. Equilibrium points in nonzero-sum n-person submodular games. *Siam Journal on control and optimization* **17**: 773–787.

Varin C, Reid N, Firth D. 2011. An overview of composite likelihood methods. *Statistica Sinica* : 5–42.

Wooldridge JM. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, second edition edition.

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