UNIVERSITY OF CALIFORNIA,
IRVINE

Essays on Network Economics

DISSERTATION

Submitted in Partial Satisfaction of the Requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Hanqiao Zhang

Dissertation Committee:
Professor Matthew Harding, Chair
Associate Professor Ying-Ying Lee
Associate Professor Jiawei Chen

2023

DEDICATION

This dissertation is dedicated to all those who believed in me when I wavered in believing in myself. Your faith, encouragement, and patience have been the rock upon which this work stands.

To my advisor, whose insightful guidance and rigorous academic standards have not only shaped this dissertation but also profoundly impacted my personal and academic growth. Your mentorship has been a crucial pillar in the edifice of this dissertation.

To my parents, whose love and sacrifices have been the beacon guiding me through the trials. Your enduring belief in the potential that lies within me has been my driving force.

To my partner, whose understanding, patience, and support have been priceless. The hours spent listening, reassuring, and simply being there have rendered this adventure enjoyable.

Lastly, to all my friends and peers who have shared this journey with me, your camaraderie has been a source of strength and comfort.

The completion of this work is not my achievement alone, but rather a testament to the collective effort of all those who have supported me along the way.

Dedicated with love and gratitude.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am profoundly grateful to a number of individuals whose support and guidance were pivotal in the completion of this dissertation.

First and foremost, I would like to express my sincere gratitude to my advisor and committee members. Professor Matthew Harding, who, as my Chair, has provided invaluable guidance and mentorship throughout my research journey; Professor Ying-Ying Lee and Professor Jiawei Chen, whose expertise and insightful advice greatly enhanced the quality of this work.

I would like to extend my appreciation to Professor Ivan Jeliazkov, Professor Yingying Dong, and Professor Jan K. Brueckner for their insightful advice and constructive criticism throughout my research. Their encouragement and wisdom have been absolutely essential.

I would like to acknowledge Professor Samantha Vortherms, who not only contributed to the dataset construction for the firm network chapter, but also shared her profound ideas that significantly enriched this work.

My heartfelt thanks go to all my peers and classmates, especially Hao-Che (Howard) Hsu, Yu Lei, Joy D. Xiuyao Yang, Nils Breitmar, and Jieyu Gao. Their camaraderie, support, and thoughtful input have made this journey both enjoyable and rewarding.

Lastly, I would like to express my deepest gratitude to my parents, whose love, encouragement, and faith in my abilities have been my greatest source of strength. My special thanks go to my partner, Jiawei Wang, whose support and understanding have been instrumental throughout this journey.

To all those mentioned and many others who have contributed to this journey, thank you. Your collective support has made this dissertation possible.

VITA

Hanqiao Zhang

EDUCATION

Bachelor of Science in Finance	**2016**
Beijing University of Technology	*Beijing, China*
Masters of Science in Applied Economics	**2018**
Johns Hopkins University	*Washington D.C.*
Masters of Art in Economics	**2020**
University of California, Irvine	*Irvine, CA*
Doctor of Philosophy in Economics	**2023**
University of California, Irvine	*Irvine, CA*

FIELDS OF STUDY

Econometrics, Industrial Organization, Machine Learning

ABSTRACT OF THE DISSERTATION

Essays on Network Economics

By

Hanqiao Zhang

Doctor of Philosophy in Economics

University of California, Irvine, 2023

Professor Matthew Harding, Chair

The dissertation comprises three empirical papers, each analyzing a novel network-structured dataset across three different contexts: enterprises, social networks, and academic collaborations.

The first study uncovers the largely unexplored area of enterprise exits in the context of Shenzhen's electronics manufacturing industry. It reveals the spillover effects of such exits on neighboring firms, with significant neighborhood effects found at the industry group level but not the industry class level.

The second paper challenges the "Death of Distance" proposition by examining how geographical proximity continues to influence online social networks. It deciphers the complex interplay between physical distances and users' online behaviors, presenting country-specific patterns in how distance affects the likelihood of link formation. Particularly, proximity dependence appears to be stronger for potential in-person connections and weaker for strong social ties.

The final study capitalizes on the disruptive force of the COVID-19 pandemic on academic collaboration. It elucidates how peer effects and co-authorship dynamics influence the productivity of economics scholars. The peer effect is significant in the pre-pandemic period but not during the pandemic period, enhancing the understanding of how research collaborations shape knowledge production.

Chapter 1

Exit Spillovers of Foreign-invested Enterprises in Shenzhen's Electronics Manufacturing Industry

Neighborhood characteristics have been broadly studied with different firm behaviors, e.g. birth, entry, expansion, and survival, except for firm exit. Using a novel dataset of foreign-invested enterprises operating in Shenzhen's electronics manufacturing industry from 2017 to 2021, I investigate the spillover effects of firm exits on other firms in the vicinity, from both the industry group and the industry class level. Significant neighborhood effects are identified for the industry group level, but not the industry class level.

1.1 Introduction

Spatial dependence on various firm behaviors, such as birth, entry, survival, and expansion, is widely studied in the context of the influences of local characteristics, industry agglomeration, and specialization. However, few papers have associated it with firm exit. Empirical evidence shows significant geographical patterns and neighborhood effects for firm exit. (Arcuri, Brunetto and Levratto, 2019) investigates firm exits in France and discovers that places with high exit rates are more likely to be surrounded by similar ones. (Sarmiento and Wilson, 2007) concludes the spatial binary lagged dependent variable is the most powerful in explaining firm exit in the U.S. baking industry.

The contribution of this chapter is three-fold. First of all, using a new dataset of foreign-invested enterprises in Shenzhen's electronics manufacturing industry, the paper adds evidence to the existence of spatial dependence on firms' exit behaviors. Secondly, it depicts a picture of divestment in Shenzhen's manufacturing industry across time in the context of foreign-invested enterprises' exits. I analyze neighborhood effects and important factors to firm exits under different hierarchies of industry classification categories. Thirdly, it looked into how neighborhood effects of firm exit behaviors may be impacted by large external shocks. During the past 5 years, at least two external shocks impacted all foreign-invested enterprises that were operating in mainland China: raised tariffs and political risks brought by the 2018 U.S.-China trade war, and rapid changes in the business environment induced by COVID-19. The shocks not only affect firms' probability to exit directly, but they may also generate non-trivial neighborhood effects. For example, tariffs imposed on a specific industry result in the leaving of some firms from highly specialized production geography, and these firm exits may increase or decrease the likelihood of other firms' leaving due to the loss of benefits from specialization, disruption of the supply chain, or the growth of market share freed by former competitors.

The second section reviews the literature that is closely related to the spillover effects of foreign-invested enterprise exits. The third section theorizes the spillover effects on firm exits. The fourth section introduces the novel foreign-invested enterprises in China dataset. The fifth section describes the spatial lagged probit model and the applied GMM estimator. The sixth section illustrates the empirical results, and the last section concludes.

1.2 Literature Review

Firm exit is one of the most discussed topics since (Baldwin and Gorecki, 1991), and its determinants have been studied in a large body of literature, see a systematic review in (Cefis et al., 2022). Although local context has been studied extensively for firm birth (Calá, Manjón-Antolín and Arauzo-Carod, 2016, Audretsch, Dohse and Niebuhr, 2015, Lee, Hong and Sun, 2013, Levratto and Carré, 2014), firm entry (Cheratian, Goltabar and Calá, 2021), firm survival (Huiban, 2011, Craioveanu and Terrell, 2016), and firm growth (Levratto and Garsaa, 2016), it has not been related to the firms' exit behaviors until recent years (Weterings and Marsili, 2015, Ferragina and Mazzotta, 2015).

Two strands of related literature are industrial agglomeration and specialization. The former refers to the phenomenon that firms tend to cluster geographically (Audretsch and Feldman, 1996, Porter et al., 1998), and the benefits may outweigh the disadvantages brought by higher industrial densities, such as less shared resources and fiercer competition. Micro-foundations are offered by (Duranton and Puga, 2004) based on Marshall's trinity (Marshall, 2009): matching, sharing, and learning. Industrial clustering could produce better outcomes in matching employer and employee in terms of both the matching quality and probability (Rosenthal and Strange, 2001). Firms may share a wide variety of input suppliers, expensive facilities, and the gains of individual specialization (Baumgardner, 1988). The latter emphasizes the concentration of enterprises in a particular industry or sector in a given region, which could facilitate innovation (Duranton and Puga, 2001), knowledge spillovers (Jovanovic and Rob, 1989), infrastructure in the region, and accumulation. Empirical results also provide supporting evidence. (Cainelli, Montresor and Vittucci Marzetti, 2014) found that specialization decreased firm exit rates in the short run, especially for the low-tech industry. In the case of (Power, Doran and Ryan, 2019), specialization reduces exits at the firm level but not the regional level. In the context of China, (Fan and Scott, 2003) substantiates positive relationships between spatial agglomeration and productivity in various Chinese manufacturing sectors. To my best knowledge, the only paper that directly estimates spatial correlations of firm exit is (Arcuri, Brunetto and Levratto, 2019), in which both the dependent variable, exit rate, and control variables are at the aggregated French department level. They identify significant positive spatial autocorrelation of firm exits. Compared to the previous study, this paper models the spatial dependence of enterprises' exit behaviors using firm-level observation and control variables.

Focusing on the closure of foreign-invested enterprises, this paper also fits into the fast-growing foreign divestment literature. Compared to foreign direct investment, comparatively little attention has been paid to this area due to data limitations. Observations of enterprises' divestment not only require panel data (Lee and Madhavan, 2010), but also need the enterprises to willingly share divestment decisions that may show business failures (Benito, 1997). Some theories that explain foreign subsidiary divestment highlight the importance of local market conditions and interconnections among enterprises. From the resource-based point of view, (Barney, 1991) identifies four empirical indicators that determine a foreign subsidiary's potential to generate competitive advantage and thus could protect it from divestment: value, rareness, imitability, and substitutability. Eclectic paradigm, developed by (Dunning, 1980), considers three factors – ownership, location, and internalization – that influence the decision-making of multinational enterprises when engaging in foreign divestment decisions. When an enterprise divests from a region, it may signal a decline in the value of its advantages, prompting other enterprises with similar resources to reconsider their presence in the region. Unfortunately, few empirical papers have yet been found to focus on the relationship among enterprises' divestment decisions in the same region, or how remaining enterprises may act in response to the divestment of other enterprises in the vicinity.

1.3 Theorizing Neighborhood Effects on Enterprise Exits

Exit decisions of enterprises in the neighborhood may not be made independently. Research in industrial agglomeration and specialization shows that enterprises tend to locate near others that produce either homogeneous or similar products of different magnitudes due to the benefit brought by spillover effects or economies of scale. If regional enterprise density declines due to the exit of neighbors, an enterprise may no longer enjoy the positive externalities, such as knowledge spillovers, access to suppliers and customers, larger labor pool, etc., and decide to exit itself.

Neighborhood effects on exit decisions could vary by the business relationships among the enterprises in the same industry. On one hand, the exits of direct competitors could create a void that allows the remaining firms to capture a larger market share. If more competitors are leaving, the overall competitiveness of the regional market may even be impacted. It becomes easier for the remaining enterprises to collude and increase their market power, making them less likely to exit the market. On the other hand, the exit of an enterprise in a neighborhood could disrupt the local supply chains that other parties in the area have been relying on. For example, if the exiting enterprise was a supplier or customer of other enterprises in the area, the remaining enterprise may also consider leaving because it struggles to find alternative suppliers or customers, and thus may have higher costs and lower profitability. The business relationships among enterprises in the same industry may not be identical according to the definition of the industry, or its aggregate level in the data. The lower the aggregate level, the more subdivided industry the enterprises may operate within, thus having a stronger intensity of competition. Higher industry aggregate level may bring in more indirect competitors who produce non-homogeneous goods but satisfy the same general demand, or even support enterprises that operate in the upstream or downstream subdivided industry. In this case, the competition intensity within the same industry would be lower, and connections among enterprises would be more complicated.

In the presence of neighborhood effects, the U.S.-China trade war in 2018 and 2019 could be a significant disruptor to regional industrial clusters beside the tariff impositions. As some enterprises exit the market due to higher tariffs, it would be more difficult for the remaining firms to maintain economies of scale, and keep lower costs of sourcing raw materials or components. A downward spiral of economic activity in the surrounding area may occur, e.g. demand for related services like logistics and transport is reduced. This could further weaken the economic viability of the staying firms, contributing to their probability of exiting and creating a domino effect.

Apart from the spillover effects generated by nearby enterprises, the exit decision could be affected by a series of enterprise-level variables as well. Firstly, more productive and efficient enterprises are less likely to exit the market because they could better compete with other enterprises in the market by producing at a lower cost, offering higher quality products or services, and investing more in research and development. Scholars found that higher firm productivity, both in technical and labor efficiency, is related to higher survival time and a lower rate of market exits (Muzi et al., 2022), (Aga and Francis, 2017). Secondly, the way firms are structured and governed may play an important role in their exit decisions. A sole proprietorship or partnership may be more likely to exit the market compared to a corporation because the liabilities of the former are tied to the personal assets of the owner, whereas the liabilities of the latter are limited to the assets of the corporation. It could make it riskier for small enterprises to stay in the market if they face financial difficulties or other challenges. There is also some relevant empirical evidence. (Cotei and Farhat, 2018) finds firm's legal structure, such as operating as a corporation or sole proprietorship, affects its acquisition outcome through innovation and employment growth. (Goktan, Kieschnick and Moussawi, 2018) argues that more than many economic factors, some corporate governance features, such as the size of independent boards, and restrictions on shareholder governance, are more important in determining if a company would exit by M&A, going private, or going bankrupt. Thirdly, enterprises' ages and sizes could be related to their probability to exit. On one hand, younger and smaller enterprises may be more likely to exit the market, because they are typically less established and have less experience. They may face greater challenges in accessing resources, such as capital and customers, or be more vulnerable to changes in the market or competition. On the other hand, older enterprises may exit if they are less flexible and adaptable compared to younger ones, making them difficult to respond to changes or adapt to new technologies. Research also shows that foreign enterprises become less attached to local markets as they age, grow "footloose" and have higher market exit rates, see (Mata and Freitas, 2012), (Coucke and Sleuwaegen, 2008).

1.4 Foreign Electronics Manufacturer in Shenzhen

The dataset is constructed with two sources: the foreign-invested enterprises in China (FIEC) database provided by (Vortherms and Zhang, 2021), and Qichacha, a website that delivers business data on China-based private and public companies. The FIEC dataset covers officially-registered multinational corporations, from 2014 to 2019, that operate in Mainland China and have at least one foreign investor. Reports are collected annually from the website of the Chinese Ministry of Commerce, which requires every foreign-invested enterprise to register its information before June. This includes name in both English and Chinese, date of registration, date of establishment, industry, business practice, firm type, address, registered capital, realized capital, area of registration, investors, and annual reports. Qichacha is an information query platform that collects and collates public information from government websites, such as the National Enterprise Credit Information Publicity System, China Executive Information Disclosure network, etc. I combine the FIEC dataset with the lists of enterprises invested by foreign investors and investors from Hong Kong, Macao, and Taiwan exported from Qichacha as of January 2023. This includes filling in missing values in the FIEC dataset and updating outdated registration information, such as business scope, address, and annual reports. The industry descriptions in the FIEC database also enable the industry classification of foreign-invested enterprises on four different aggregate levels based on the Chinese standard for industrial division[1]. The standard classifies enterprises based on the productive activities they engage, thus enterprises that are grouped together primarily perform the same type of production. The following figure is a minimal example of the four industry aggregate levels in the context of manufacturing.

Firms are more homogeneous in terms of the productive activities that they engage in if the classification category that they belong to is more narrow or specialized. For example, if both firms from the same city specialize in computer parts manufacturing, they may cater to the same group of customers, such as original equipment manufacturers, computer repair and upgrade shops, etc.

[1]This refers to the "GB/T 4754—2017 Industrial Classification for National Economic Activities".

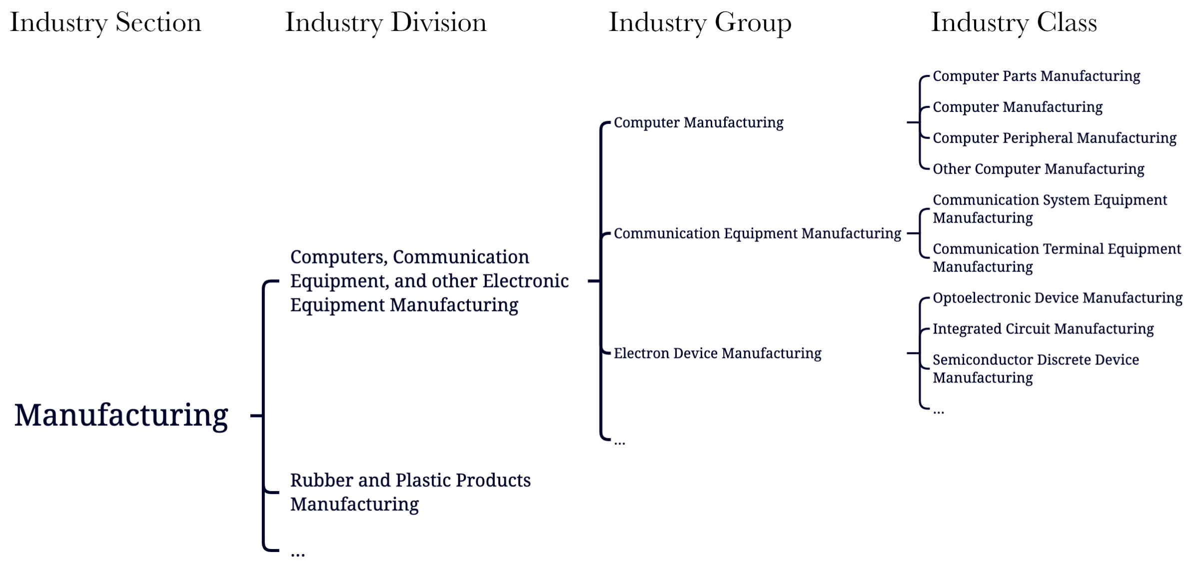


Figure 1.1: Four Levels of Classification Categories for the Manufacturing Industry-section

However, firms are more diversified if we move to broader levels of classification categories. At the industry group level, firms could still operate in the same industry class, or they could be in the supporting industries of another: computer parts manufacturers may be the local suppliers of computer manufacturers.

In the present study, we focus on the enterprises that operate in the electronics manufacturing industry in Shenzhen, 2017, for a few reasons. Firstly, as a major hub for electronics manufacturing in Guangdong province, Shenzhen has a high density of industrial parks and special economic zones. This unique landscape facilitates a significant concentration of foreign-invested enterprises in the region. Secondly, the electronics manufacturing industry has a complex and extensive supply chain, with multiple levels of suppliers and subcontractors. This makes the industry more susceptible to spillover effects, as the exit of a key foreign-invested enterprise could disrupt the supply chain and create uncertainties for other enterprises. Thirdly, in the background of the U.S.-China trade war starting in 2018, the electronics manufacturing industry faces increasing pressure due to geopolitical tensions and trade disputes, and the spillover effect may bring collateral damage that leads to an additional number of firm exits. These contexts make the industry more relevant to study the neighborhood effects of foreign-invested enterprises' exits.

Prior to 2019, the Chinese Ministry of Commerce website provides a full list of foreign-invested enterprises that are operating in mainland China in the current year, so the exit time of each enterprise is marked as the year at which it no longer shows up in the list. Starting in 2020, the website realizes a technology upgrade and displays the search result only for the typed-in enterprise. Therefore, the exit year of each enterprise that remains active through 2019 is marked as the last year that it has submitted its annual report that could be verified on Qichacha. Depending on the identity of foreign investors, the exit behavior could be the subsidiary closure of a foreign multinational corporation, or the abandonment of the Chinese market by a wealthy foreign individual.

For enterprises that operate as early as 2017 in Shenzhen, I report the numbers and percentages of enterprise exit each year from the computers, communication, and other electronic equipment manufacturing division, the broader manufacturing section, and all industry sections, in the following table. For a broad overview of enterprises' exits in all Chinese first-tier cities from 2014 to 2021, see A.1.

Table 1.1: Foreign-invested Enterprises' Yearly Exits in Shenzhen, 2017-2021

	Computers, Communication & Other Electronic Equipment Manufacturing (Division)		Manufacturing (Section)		All Industry Sections	
	Enterprises	Exits (%)	Enterprises	Exits (%)	Enterprises	Exits (%)
2017-2018	1,458	92 (5.94%)	8,097	441 (5.17%)	46,889	1,727 (3.55%)
2018-2019	1,300	158 (10.84%)	7,301	796 (9.83%)	42,270	4,619 (9.85%)
2019-2020	1,170	130 (10.00%)	6,434	867 (11.88%)	38,109	4,161 (9.84%)
2020-2021	1,116	54 (4.62%)	6,109	325 (5.05%)	36,116	1,993 (5.23%)

Following the outbreak of the trade war in 2018 and the pandemic in 2019, the exit rates in these two years are both at the highest level. The electronics manufacturing industry and the broader manufacturing industry have higher exit rates than the overall industries in general.

Summary statistics of the firm-level characteristics are shown in the table. Distributions of capital size and aggregate investment of the enterprises have a mass below 1,000 and long right tails. 81.16% of the enterprises are foreign-owned, 14.19% are joint-venture companies, and 4.65% have other types, such as contractual joint venture, share-holding, and partnership. For joint ventures, the median percentages of the foreign contribution of both the registered capital and realized capital in joint ventures are about 45%. As for places of origin, most of the foreign-invested enterprises are registered in Asia. More than half of the enterprises (64.52%) are registered in Hong Kong, followed by Taiwan (5.35%), British Virgin Islands (5.03%), Samoa (3.81%), and the United States (2.32%). Following (Vortherms and Zhang, 2021), I also include rough indicators of tariff exposure during the 2018 trade war and enterprises' importer/exporter status. The former marks enterprises that operate within tariffed industry class, using tariff data from (Bown, 2019), and the latter filters if enterprises mention the keywords "export" or "import" in their business scopes. More than 94% of the enterprises are exposed to tariffs according to the industry class, and at least 51% engages in import and export trade business.

Table 1.2: Summary Statistics of Categorical Variables

	Shenzhen
Foreign-owned	81.16%
Joint-venture	14.19%
U.S. Registered	2.32%
Tariffed Industry	94.39%
Importer/Exporter	51.55%

Table 1.3: Summary Statistics of Numerical Variables

	Mean	Median	St.Dev.
Aggregate Investment	13,302.90	760.00	172,804.50
Registered Capital	6348.06	547.00	64785.70
Foreign Contribution(%) for Joint-venture	47.83	45.00	0.24
Realized Capital	1000.00	103.10	7748.32
Foreign Contribution(%) for Joint-venture	47.05	45.71	0.25

Note: All currencies are in $10,000

1.5 Spatial Lagged Probit Model

To estimate the neighborhood effects on enterprise exits, I apply the classic spatial lagged probit model in spatial econometrics literature (Anselin, 1988, Anselin, Florax and Rey, 2013, LeSage and Pace, 2009). There are n foreign-invested enterprises in Shenzhen's electronic manufacturing industry that remain active in a given year. The observed dichotomous exit decision, Y_i, for enterprise $i = 1, 2, \cdots, n$, depends on the value of a latent continuous variable, interpreted as its propensity to exit, Y_i^*:

$$Y_i = \begin{cases} 1, & Y_i^* \geq 0 \\ 0, & Y_i^* < 0 \end{cases}$$

Enterprises' latent propensity to exit in vector form, $\boldsymbol{Y}^*$, is assumed to be the linear combination of itself and the matrix of firm-level control variables, $\boldsymbol{X}$:

$$\begin{aligned} \boldsymbol{Y}^* &= \boldsymbol{\rho} \boldsymbol{W} \boldsymbol{Y}^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \text{MVN}\left(\boldsymbol{0},\ \sigma_{\boldsymbol{\epsilon}}^2 \boldsymbol{I} \right) \end{aligned}$$

where $\boldsymbol{\epsilon}$ is assumed to follow a multivariate normal distribution, and $\sigma_{\boldsymbol{\epsilon}}^2$ is the variance of the error term. $\boldsymbol{X}$ includes enterprises' years of operation, the size of registered capital, the percentage of registered capital contributed by foreign investors, regions of registration, legal form, the tariff indicator, and the importer/exporter indicator. $\boldsymbol{W}$ is an exogenously specified weight matrix with 0s on the diagonal. I define elements in $\boldsymbol{W} = [W_{ij}]$ as the inverse geographical distance of enterprise i and j:

$$W_{ij} = \begin{cases} \dfrac{1}{d_{ij}}, & \text{if } i \text{ and } j \text{ belong to the same industry} \\ 0, & \text{if } i \text{ and } j \text{ are from different industries} \end{cases}$$

where d_{ij} is the great-circle distance computed from the two enterprises' longitudes and latitudes. $\boldsymbol{W}$ is then row-normalized as $W_{ij}/\sum_j W_{ij}$, so that the enterprise's propensity to exit is a weighted average of neighboring enterprises' propensities to exit, excluding itself. Geographically closer enterprises generate stronger effects by assumption. Four models are estimated for a given year using four $\boldsymbol{W}$ that allow for the correlation of enterprises' propensities to exit within four aggregate levels of the industry respectively.

It was shown that spatial dependence could only be introduced through the latent variable, e.g. models like $\boldsymbol{Y}^* = \boldsymbol{\rho W Y} + \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, or $\boldsymbol{Y} = \boldsymbol{\rho W Y} + \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ are not algebraically consistent, see (Beron and Vijverberg, 2004, Klier and McMillen, 2008). In our case, the spatial lagged model assumes the propensity of each firm to exit the market depends on other nearby firms' propensities to exit, instead of whether other nearby firms have actually left. The application is appropriate because firms' market exit decisions, in the language of (Klier and McMillen, 2008), are "forward-looking in nature". For example, when an enterprise decides whether to exit a specialized town, it may expect that neighboring enterprises are reluctant to leave and give up the benefits brought by agglomeration. The enterprise anticipates the low values of $\boldsymbol{Y}^*$ could hold the cluster or even attract more firms, which further reduces fixed costs or generates greater competitive advantages.

The reduced-form equation of the spatial lagged probit model could be derived:

$$\boldsymbol{Y}_{\boldsymbol{i}}^* = \boldsymbol{X}^* \boldsymbol{\beta} + \boldsymbol{u}$$

where $\boldsymbol{X}^* = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{X}$, $\boldsymbol{u} \sim \text{MVN}(0, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \left[(\boldsymbol{I} - \rho \boldsymbol{W})' (\boldsymbol{I} - \rho \boldsymbol{W})\right]^{-1}$.

The model leads to inconsistent and inefficient estimates due to heteroskedastic errors. Many efforts were made to overcome the problems induced by spatial dependence from the aspects of both theoretical and empirical (Fleming, 2004), and a few estimators were proposed. From the Bayesian's perspective, (McMillen, 1992) proposes an EM algorithm that replaces the latent $\boldsymbol{Y}^*$ with expected values in the E-step, then estimates model parameters with maximum likelihood in the M-step. (LeSage, 2000) suggests a Gibbs sampler that produces random draws of y_i^* from a multivariate truncated normal distribution conditional on all other model parameters. (Beron and Vijverberg, 2004) comes up with a recursive importance sampling (RIS) algorithm that directly evaluates the probit likelihood function. From the frequentist's perspective, (Pinkse and Slade,

1998) proposes a generalized method of moments (GMM) estimator based on the spatial error probit model, and it was later linearized around zero interdependence in (Klier and McMillen, 2008).

The RIS and Bayesian strategies are shown to be able to provide accurate estimates for spatial lagged probit models. The GMM estimators, being instrumented-approximation methods, work well only when the samples are large and the spatial dependence is not strong. However, they are computationally much more efficient with running times orders of magnitude shorter (Calabrese and Elkink, 2014). For this reason, I apply the linearized GMM estimator developed in (Klier and McMillen, 2008) to the data, because it is the only feasible one in estimation time to run one model for each year at each classification category. Spatial dependence parameters from two randomly chosen models are estimated by the Gibbs sampler, and differences between the results and the one generated by the linearized GMM estimator are at the second digit. Nevertheless, this is still considered a limitation of the estimation strategy of this paper, and more models should be run to ensure the GMM estimates are valid.

I then illustrate the GMM estimator for the spatial lagged probit model in detail. Its log-likelihood function could be easily written from the reduced-form equation as:

$$l\left(\boldsymbol{\beta},\rho|\boldsymbol{X},\boldsymbol{W},\boldsymbol{Y}^{*}\right) = \sum_{i=1}^{n}\left\{ y_i \cdot lnP\left(y_i=1|\boldsymbol{x_i},W_{ij},y_j^*\right)+(1-y_i)\cdot ln\left[1-P\left(y_i=1|\boldsymbol{x_i},W_{ij},y_j^*\right)\right]\right\}$$

$$P\left(y_i=1|\boldsymbol{x_i},W_{ij},y_j^*\right) = P\left(y_i^*\geq 0|\boldsymbol{x_i},W_{ij},y_j^*\right)=P\left(\boldsymbol{x_i^{*\prime}\beta}+u_i\geq 0|\boldsymbol{x_i},W_{ij},y_j^*\right)=\Phi\left(\frac{\boldsymbol{x_i^{*\prime}\beta}}{\sigma_i}\right)$$

where $\Phi(\cdot)$ is the CDF of standard normal distribution, σ_i is the ith standard deviation according to $\boldsymbol{\Sigma}$. (Pinkse and Slade, 1998) derived the sample moment condition that accounts for the heteroscedastic errors:

$$m\left(\beta,\rho\right) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z_i'}\left\{\frac{\left[y_i-\Phi\left(\frac{\boldsymbol{x_i^{*\prime}\beta}}{\sigma_i}\right)\right]\phi\left(\frac{\boldsymbol{x_i^{*\prime}\beta}}{\sigma_i}\right)}{\Phi\left(\frac{\boldsymbol{x_i^{*\prime}\beta}}{\sigma_i}\right)\left[1-\Phi\left(\frac{\boldsymbol{x_i^{*\prime}\beta}}{\sigma_i}\right)\right]}\right\}$$

where $\boldsymbol{z_i}$ is the ith row of $\boldsymbol{Z}$, the matrix of instruments that are composed of the control variables $\boldsymbol{X}$. The parameters of interest satisfy the moment condition $m(\beta,\rho)=0$. When the number of moment conditions exceeds the number of unknown parameters, the model could be estimated by:

$$\text{argmin}_{\beta,\rho\in\Theta} \quad \boldsymbol{m}\left(\beta,\rho\right)' \boldsymbol{M}\boldsymbol{m}\left(\beta,\rho\right)$$

where Θ is the parameter space, $\boldsymbol{M}$ is a positive-definite matrix that assigns weights to different moment conditions. This equation does not have an analytical solution and needs to be solved with non-linear optimization algorithms. The computation could be burdensome because each iteration involves the inverse of a n by n matrix, $\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}$, while evaluating any candidate value of ρ. If M is specified as $\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}$, the estimator becomes non-linear two-stage least squares ((Amemiya, 1975), (Amemiya, 1974)) with the objective function:

$$\text{argmin}_{\beta,\rho\in\Theta} \quad \hat{\boldsymbol{e}}\left(\beta,\rho\right)' \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\hat{\boldsymbol{e}}\left(\beta,\rho\right)$$

where $\hat{\boldsymbol{e}}\left(\beta,\rho\right)$ is the generalized probit residuals inside the curly brackets of $m\left(\beta,\rho\right)$. To estimate the parameters, we first assume the initial value $\boldsymbol{\Gamma}_0 = \left(\rho,\boldsymbol{\beta}\right)'$, then compute $\hat{\boldsymbol{e}}_0\left(\beta,\rho\right)$ and gradient terms $\boldsymbol{G} = \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{\Gamma}}$. Regress $\boldsymbol{G}$ on $\boldsymbol{Z}$ to obtain the predicted value $\hat{\boldsymbol{G}}$, and the new estimates $\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_0 + \left(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}}\right)^{-1}\hat{\boldsymbol{G}}'\hat{\boldsymbol{e}}_0\left(\beta,\rho\right)$. This process is iterated until convergence. This GMM estimator is computationally challenging because the gradient term, $\frac{\partial \boldsymbol{P}}{\partial \rho}$, contains $\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}$, and each iteration involves the inversion of an n-dimensional matrix. To circumvent the problem, (Klier and McMillen, 2008) linearized the estimator around the convenient starting point of standard Probit model, thus greatly simplified the gradient terms. In particular, they first estimate $\hat{\beta}_0$ with standard Probit model, compute the generalized error term $u_0 = y_i - P\left(y_i = 1\right)$ and the gradient terms G_{β_i} and G_{ρ_i}. In the second step, they regress G_{β_i} and G_{ρ_i} on Z to obtain the predictions $\hat{G}_{\beta_i}$ and $\hat{G}_{\rho_i}$. Then they regress $u_0 + G'_{\beta}\hat{\beta}_0$ on $\hat{G}_{\beta_i}$ and $\hat{G}_{\rho_i}$ for the estimated values of β and ρ.

1.6 Empirical Results

I focus on enterprises that operate in Shenzhen's computers, communications, and other electronic equipment manufacturing industry division in 2017, then estimate the spatial lagged Probit model for 2018, 2019, 2020, and 2021, at the industry group and industry class level respectively. In each model, I alter the specification of W so that only spatial dependence among enterprises within the same classification category is allowed. For example, at the industry class level, I assume that the propensities to exit are correlated among computer parts manufacturers, but not between computer parts manufacturers and computer peripheral manufacturers. Enterprises within the same industry class produce homogeneous products. The industry group level also includes enterprises that operate in the upstream, downstream, or supporting industries of the business. Although it would be an interesting exploration, I have not estimated the model at the broader industry division and industry section level. This helps filter certain industry-specific characteristics, for example, while considering whether to exit the market, a computer parts manufacturer may not take into account the tendencies of nearby plastic product manufacturers.

Table 1.4: Neighborhood Effect on Enterprise Exits in Shenzhen's Manufacturing of Computers, Communications and other Electronic Equipment

	2017-2018	2018-2019	2019-2020	2020-2021
Industry Group	0.77^{***}	1.00^{***}	1.65^{***}	2.28^{***}
	(0.25)	(0.38)	(0.43)	(0.43)
Industry Class	-0.22	-0.001	0.55	0.94
	(0.18)	(0.41)	(0.53)	(0.62)
Note:		*p<0.1; **p<0.05; ***p<0.01		

Firstly, significant spillover effects are identified at the industry group level, but not at the industry class level. This implies that enterprises' propensities to exit may not be affected by their competitors, but are significantly influenced by enterprises that operate in their upstream or downstream business.

Secondly, at the industry group level, the spillover effect gradually becomes larger from 2018 to 2021. This could be explained by supply chain disruptions. The trade war leads to increased tariffs and restrictions on trade, which disrupts the supply chains. This could cause some foreign-invested enterprises to exit the market or reduce their operations, affecting the supply chains of other enterprises in the same industry. The COVID-19 pandemic further exacerbated supply chain disruptions due to lockdowns and restrictions on the movement of goods and people. These combined effects lead to an increased spillover effect, as more enterprises may consider exiting the market or scaling down their operations due to uncertainties and disruptions in the supply chain. In addition, both the trade war and the COVID-19 pandemic created significant uncertainty in the global market. This could lead to reduced foreign investment in China and increased risk aversion among foreign-invested enterprises. As a result, some enterprises might exit the market, leading to a larger spillover effect on the remaining enterprises in the same industry.

Table 1.5: Covariate Effects on Enterprise Exits in Shenzhen's Manufacturing of Computers, Communications and other Electronic Equipment

	2017-18		2018-19		2019-20		2020-21	
	Group	Class	Group	Class	Group	Class	Group	Class
Years of Operation	-0.13^{***}	-0.13^{***}	-0.08^{**}	-0.08^{**}	0.03	0.04	0.02	0.04
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Registered Capital	0.21	0.21	0.09	0.10	0.23	0.28^{*}	0.21	0.24^{*}
	(0.24)	(0.24)	(0.23)	(0.23)	(0.16)	(0.15)	(0.13)	(0.12)
Foreign Contributed	0.12	0.11	−0.29	−0.30	−0.25	−0.29	−0.06	−0.11
	(0.26)	(0.26)	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)
Registration (Taiwan)	0.19	0.21^{*}	0.34^{**}	0.33^{**}	0.27^{*}	0.27^{*}	0.19	0.19
	(0.12)	(0.12)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
Registration (U.S.)	−0.03	−0.03	−0.07	−0.08	−0.07	−0.07	0.001	−0.0004
	(0.19)	(0.19)	(0.20)	(0.20)	(0.21)	(0.21)	(0.21)	(0.22)
Legal Form (Joint-venture)	0.30^{*}	0.27^{*}	0.11	0.11	−0.04	−0.06	0.10	0.07
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Is Tariffed	0.25^{**}	−0.36	0.05	−0.22	0.15	0.31	0.21	0.56
	(0.11)	(0.32)	(0.15)	(0.43)	(0.14)	(0.40)	(0.13)	(0.39)
Importer/Exporter	-0.22^{***}	-0.23^{***}	−0.10	−0.10	−0.04	−0.04	−0.03	−0.04
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)

Note: *p<0.1; **p<0.05; ***p<0.01

For the covariate effects, years of operation show a significant negative effect on market exits in 2017-18 and 2018-19. This suggests that younger firms are generally more likely to exit the market, but the effect may have lessened during 2019-20 and 2020-21. The size of the enterprises, as measured by registered capital, and the percentage of capital contributed by foreign investors, do not exhibit significant impacts on market exits across the years. Enterprises registered in Taiwan show a higher likelihood of exiting the market from 2017-2020 compared to those registered in Hong Kong. Enterprises registered in the U.S. have not shown any significant differences in market exit propensity. Joint ventures demonstrated a higher likelihood of exiting the market in 2018, but this effect is not significant in later years. Firms exposed to tariffs have a higher likelihood of market exits in 2018 at the industry group level, but no significant effects are observed in subsequent years or at the industry class level. Surprisingly, importers and exporters exhibit a negative and significant effect on market exits in 2018. This may be due to the time point that the trade war has not fully impacted the market. In later years, this effect was not significant, suggesting that the relationship between import/export status and market exits may have weakened over time. In addition, factors that are not captured in the model may also be at play. For instance, some importers and exporters might have been able to adapt to the changing trade environment by diversifying their markets, sourcing from alternative suppliers, or passing on the increased costs to customers. Government interventions, such as subsidies or other supportive policies, might have helped some affected enterprises weather the impact of the trade war, leading to an insignificant or mixed effect of importer and exporter status on market exits during this period.

1.7 Conclusion

This paper looks into how its neighboring firms may react when an electronic manufacturer in Shenzhen tends to exit in 2017-21. Using the spatial lagged Probit model, significant spillover effects are reported for enterprises that operate in the upstream or downstream industries, but not in the same industry. This indicates that the exit of an enterprise may change the dynamics of the supply chain, leading to positive spillover effects for firms in related industries. For example, the exit of a computer parts manufacturer could lead to an increase in demand for the parts of computer manufacturers. The model also captures the raising uncertainty and volatility brought by the pandemic and the U.S.-China trade war. These external shocks cause supply chain disruptions in Shenzhen's electronic industry, forcing enterprises to find alternative suppliers or change their production processes, potentially leading to more dependencies between firms in different industries. The age, size, registration area, and legal form also affect the firm's probability to exit.

Although this chapter provides the first empirical result of the spillover effects of Shenzhen electronic manufacturers' exit behaviors based on firm-level data, it has a few limitations. Firstly, I focus on 2017-2021, a time frame characterized by unique and dramatic shifts in global trade and public health. While the data provides a rich context for studying firm exit and spillover effects, it may limit the generalizability of these findings to other periods. Secondly, the sample is comprised of enterprises that have at least one foreign investor, and no local enterprises in Shenzhen are included. I alleviate the influence by focusing on the electronic manufacturing industry in Shenzhen, but the neighborhood effect may be better estimated by constructing a more comprehensive sample in the future or investigating other industries, such as the financial industry in Shanghai, and automobile manufacturers in Guangzhou. Thirdly, as is noted in the empirical result section, certain factors that influence enterprises' tendencies to exit and the spillover effects are not included in the model. For example, neighborhood effects may depend on events that happen locally, preferential policies, and macro factors, e.g. regional characteristics, industry structures, among others. Some policies may directly affect the spillover effect, e.g. China has implemented a number of policies to support industrial clustering in order to promote regional economic development and technological innovation, and improve the competitiveness of Chinese industries in the global market. Other effects are indirect, e.g. if the business environment of a city is friendly with macroeconomic stability, supporting policy and resources, then exits of enterprises in adjacent areas may generate smaller effects on one's exit decision. Some evidence is offered by previous research, e.g. (Fafchamps and Schündeln, 2013) shows that the availability of local banks helps small and medium-sized enterprises expand and acquire investment (Fafchamps and Schündeln, 2013), and it may also aid enterprises' growth and reduce bankruptcy (Arcuri and Levratto, 2020). From another angle, (Basile, Pittiglio and Reganati, 2017) discovers that local industry variety could also alleviate the exit of enterprises. Omitting these important factors could result in bias in the estimation.

Chapter 2

Analysis of Proximity Informed User Behavior in a Global Online Social Network

Despite the earlier claim of "Death of Distance", recent studies revealed that geographical proximity still greatly influences link formation in online social networks. However, it is unclear how physical distances are intertwined with users' online behaviors in a virtual world. I study the role of spatial dependence on a global online social network with a dyadic Logit model. Results show country-specific patterns for distance effect on probabilities to build connections. Effects are stronger when the possibility for two people to meet in person exists. Relative to weak ties, dependence on proximity is looser for strong social ties.

2.1 Introduction

Online social networks are a microcosm of our increasingly globalized world, connecting individuals across vast distances and cultural divides. The structure of these networks, particularly the mechanisms that drive connection and people making new friends, are areas of ongoing research interest. As we delve into the digital age, the importance of geographical proximity, a key player in traditional, offline networks, warrants close examination in the online context.

Many proposed friend recommendation algorithms for online social networks consider physical proximity as an important factor for link formation, e.g. see (Xie, 2010, Chin, Xu and Wang, 2013, Chin et al., 2012, Wang, Chin and Wang, 2011). This is based on the belief that geographical closeness can be a proxy for a higher likelihood of connection. However, other factors, such as users' ethnic group, native language, shared interests and experiences, may also dictate the preference to select whom to follow. Therefore, evaluating the significance of physical proximity in digital spaces could be very helpful in refining friend recommendation algorithms to enhance efficiency. Moreover, for global social networks, the investigation should be conducted separately with respect to users from different countries, because the importance of geographical proximity might differ due to differences in culture, regulatory environment, demographics, etc. Comparative studies have been covered in previous research, but they were typically confined to a comparison between two or three countries, relying largely on subjective self-reported data. This paper embarks on a more comprehensive exploration of the role that geographical proximity plays in the formation of digital connections, particularly in the wake of the 2020 pandemic and subsequent quarantine regulations, which catalyzed an unprecedented reliance on virtual modes of interaction.

For the rest of this paper, section two reviews related literature and generates hypothesis, section three introduces the social network dataset and the model, section four presents empirical results, and the last section concludes.

2.2 Literature Review

The study of how spatial propinquity affects the formation of social ties dates back to the 1950s, long before the advent of the Internet, in the context of face-to-face contact and phone calls. Evidence shows that physical space played an essential role and is inversely proportional to interaction frequency (Snow, Leahy and Schwab, 1981, Blake et al., 1956, Latané et al., 1995). Information technology reshapes the structure of the social network by greatly reducing the cost of making new friends and maintaining relationships, leading to the statement of "Death of Distance" in (Cairncross, 2001). It is conceivable that in the modern era, geographical location becomes less of a constraint for building connections, but there is still the desire for people to meet each other and have a cup of coffee physically. For example, (Goldenberg and Levy, 2009) argues that the ease of communication further strengthens local social ties, and the communication volume is inversely proportional to people's geographic distance. (Boase et al., 2006) finds that people who send more emails to each other also have more frequent face-to-face and phone contact. There are many other pieces of literature that demonstrate the tendency to form short-distance links or geographically closed friendship clusters (Scellato et al., 2010, Liben-Nowell et al., 2005, Hipp and Perrin, 2009, Backstrom, Sun and Marlow, 2010). It leads to my first hypothesis with respect to the influences of proximity on online link formation:

H1. The possibility of a user following another user declines along with the growth of their geographical distance.

Although online social networks exhibit distance dependence like offline, the spatial dependence may decay in different patterns or at a slower rate. This is because the cost of building online connections (clicking the "follow" button on someone's profile) is much cheaper than breaking the ice and getting acquainted with other people face-to-face. As a result, the network is usually composed of both strong and weak ties, leading to a lower strength of connections. Besides keeping in touch, people have a variety of reasons to follow other people's lives: resonate with others' thoughts, be attracted by posted pictures, or simply follow the crowd. This could be particularly true for weak ties in online directed networks, like Instagram or Twitter. Users follow others to learn new knowledge, seek values to help improve their lives or realize the lifestyles of their dreams. Thus it is possible that in the digital space, the appeal of popularity transcends geographical boundaries, leading users to form connections based on interests, influence, and cultural factors. The simplest hypothesis to test is to see whether users are more likely to follow others from popular places rather than those who are geographically close, especially when they could hardly meet in person:

H2. People are more likely to follow others from popular countries.

This could be due to several reasons. Popular countries often have a significant cultural influence that extends beyond geographical boundaries. For example, users from Italy or France may be seen as more authoritative in fields like fashion due to the perceived status of these places. Users may follow individuals from these places to keep up with trends and ideas, and have exposure to a wide array of opinions, lifestyles, and experiences, thereby enriching their own social media feeds with diverse content. Popular places may also cultivate more Internet celebrities that attract followers.

In addition, online social networks, especially global ones, comprise far more entities than offline communities from all countries and have multitudinous cultural milieus and diverse lifestyles. Research shows people from different countries have distinct motivations for using online social networks, e.g., (Kim, Sohn and Choi, 2011) found that Korean students attached more importance to gaining social support from their friends, and students from the United States put a larger weight on entertainment. Different motivations may yield divergent social behaviors. (Cardon et al., 2009) surveyed university students from 11 countries, showing that users from collectivist nations built more connections with whom they had never met. In contrast, users from individualist nations nurture more relationships offline. In other media, (Chen, Boase and Wellman, 2002) investigated people who connect with their relatives face-to-face, by telephone, and by email, reported heterogeneity within and beyond 50 kilometers in North America, other developed countries, and developing countries. These form the basis of my third hypothesis:

H3. There is heterogeneity in how users' online behaviors are affected by proximity in different regions.

Another angle that supports country heterogeneity is provided by papers that study the variation in network structures following geographical variability and unevenly distributed populations. Inhabited areas are usually rich in resources, better in climate, or have other desirable properties.

Barren lands and oceans, where there are few or no users at all, could certainly affect the estimated effect of region-specific geographical proximity, e.g., see (Butts et al., 2012).

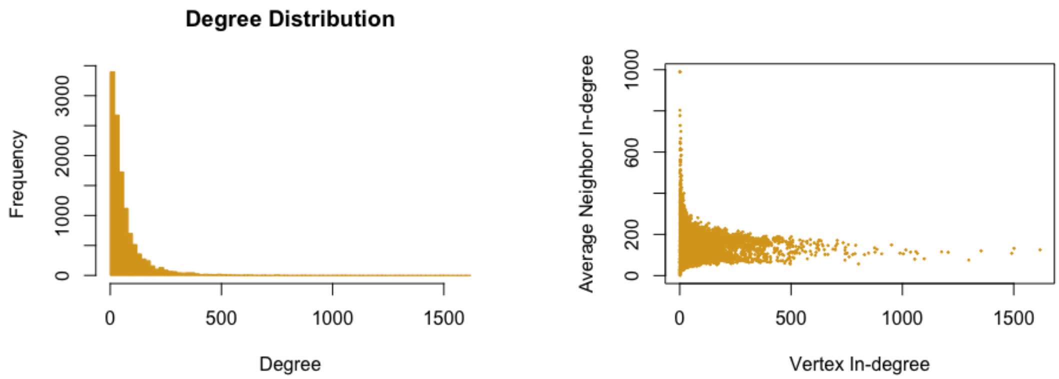
2.3 Dyadic Logit Model with Social Network Data

The social network data is collected through three processes by the mobile app: user registration, user profile, and user activity. As the first point of contact with the platform, users are asked to provide basic information about themselves during registration, such as name, email address, gender, and date of birth. In the second step, users are encouraged to complete their profiles by adding more personal information, such as their heights and weights, ethnicity, relationship status, and profile pictures. When users start to interact with the app, their online activities are logged as user behavioral data, such as the followers and followees, their liked and commented posts and stories, chat messages that are sent, etc.

I focus on a snapshot of the online social network, on December 16th, 2020, that consists of 11,992 active users, stratifed by country of origin, who have clicked the "follow" or "unfollow" button at least 4 times a day in the sample. Inactive users are not considered, because their actions are not representative of user behaviors in the network, and may obscure the relationship between geographical distance and online interactions.

Similar to other large-scale social networks in reality, the directed network is sparse with 0.59% of the total number of edges possible actually present. The in-degree distribution follows a power law of the form: $P(d) \propto cd^{-\lambda}$, where d is the vertex's in-degree, if plotted to logarithmic scale, see (Faloutsos, Faloutsos and Faloutsos, 1999, Albert, Jeong and Barabási, 1999). The power law produces a highly skewed histogram, where very few users are celebrities with plenty of followers, and most of the users are sparsely linked. The median user has 40 followers, and 0.12% of the users own 1,000 followers.

Figure 2.1: Vertex In-degree Distributions



Beyond the degree distribution, how users with different numbers of fans are linked could be seen by plotting the vertex in-degree versus the average in-degree of its neighbor. Users with few or no followers are more exploratory in their connections, as they are still trying to build their social network. They may connect with celebrity users due to their popularity or influence, while also following low-degree users who are closer to their own levels of influence. When users have more followers, their neighbors' average in-degree uncertainty decreases. This may imply that vertices with higher degrees are more selective about who they follow, and they tend to connect with other high-degree nodes, leading to a more predictable degree distribution among their neighbors. This could be due to a number of reasons, such as their desire to maintain a certain image, the limited time they have to interact with their followers, or their strategic goal to maximize their influence or reach.

I visualize the network by setting the vertices' size to correspond to the user's followers, and the vertex color to the user's country of origin. There is an edge between any two users if one user follows the other, and the colors of the edges accord with that of the followers. The user base covers all five continents and a few islands but is mainly located in Asia, Europe, and South America. For countries, 28.67% users come from Thailand, 27.65% users are from Turkey, 10.66% users are from Indonesia, 9% users come from China (Taiwan), 7.48% are from Brazil, and 5.57% users come

Figure 2.2: Social Network Visualized with Users' Geographical Locations

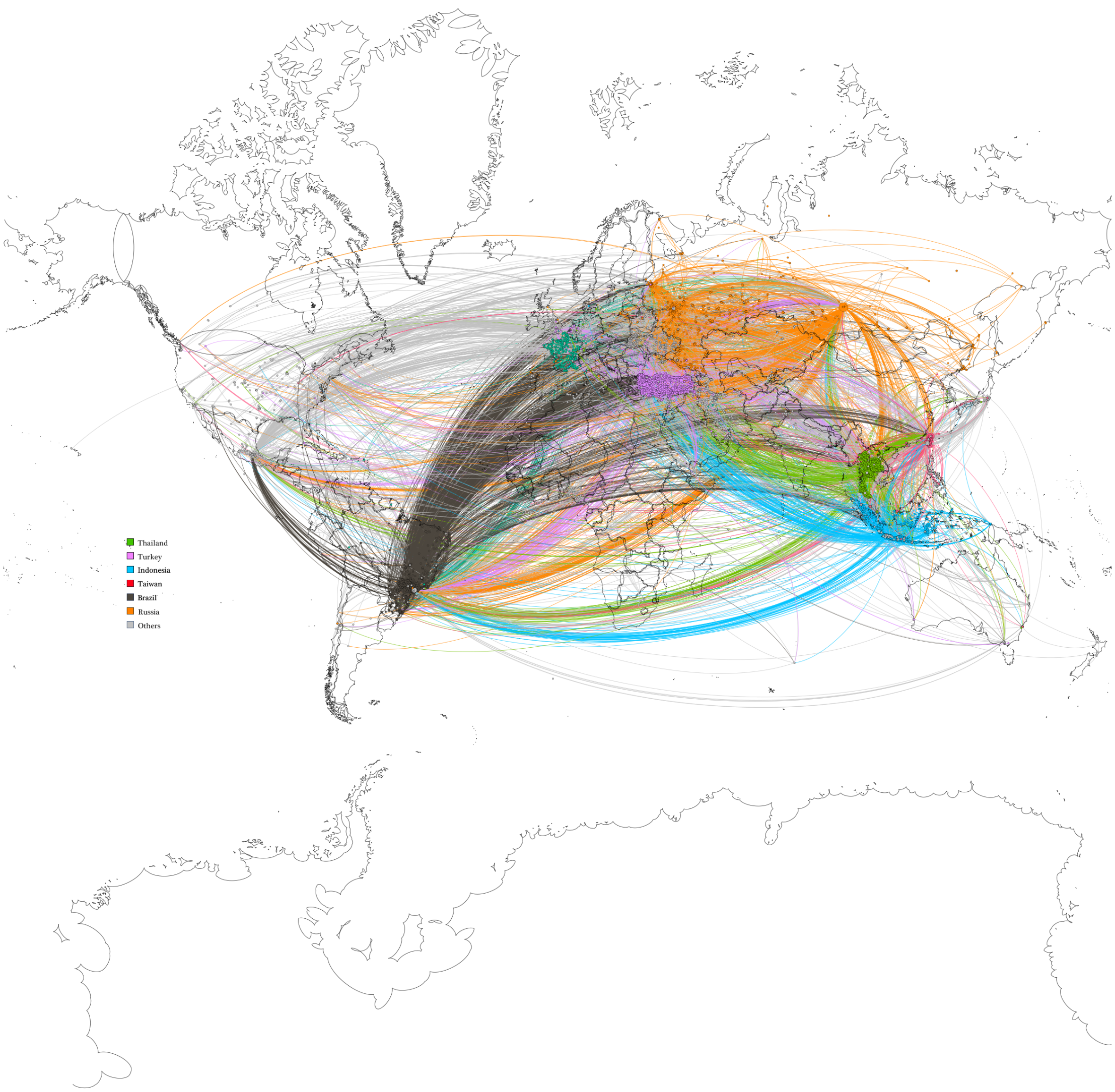
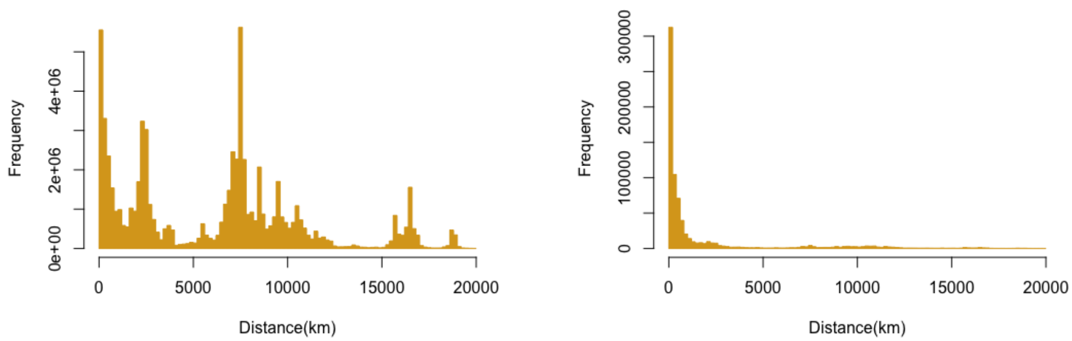


Figure 2.3: Histograms of the Distance between All Users and Mutually Connected Users

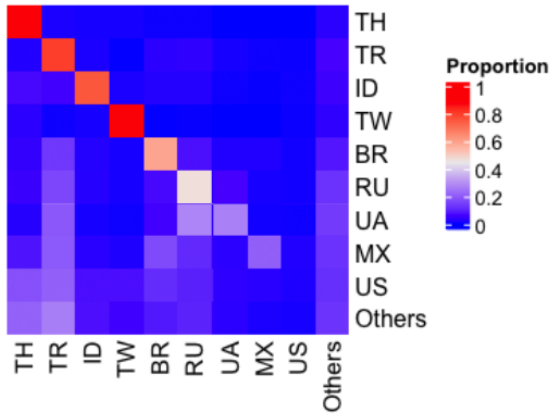


from Russia. The visualization shows that the network is truly global: people follow others from different countries or continents.

Due to the measuring scale, it is hard to see how the density of connections changes along with distance, so I compare the histogram of the distance between any two users, and the histogram of any two mutually-connected users. Distance between any two users ranges from 0 to 20,000 kilometers. The distribution has three peaks: less than 100 kilometers, near 2,500 kilometers, and around 7,000 kilometers. These may correspond to three typical scenarios: two users from the same area, two cities of the same country, and two cities of different countries. For these three groups of users, mutual connections are most easily formed for those in the first group, when two users are from the same area.

To explore the user preference for following others among different countries, I produce the heatmap, for the 10 largest countries, by listing followers' nationalities on the y-axis, and followees' nationalities on the x-axis. Users from Thailand, Turkey, Indonesia, and China (Taiwan) mainly build connections internally. Brazilian and Russian users also prefer to follow others from the same country but are open to following users from other places. More diversities are seen among people from Ukraine, Mexico, and the United States. The heterogeneity could be jointly induced by a variety of cultural, linguistic, and sociopolitical factors. For example, users are more likely to follow others who speak the same language and share the same cultural background. Higher diversity

Figure 2.4: Heatmap of Users' Countries of Following



in countries like the U.S. may be contributed by higher internet penetration, the larger size of the online community, the presence of more diaspora communities, etc.

In (Tinbergen, 1962), dyadic regression is initially used to model the logarithm of exports from country i to country j with logged Gross National Product (GNP) of both countries, logged distance of the two countries and other variables. After that, the setting is largely applied to empirical works in international trade, e.g., (Anderson, 2011, Rose, 2004). Other fields in social sciences, e.g., (Portes and Rey, 2005) explains bilateral financial assets transactions among 14 countries from 1989 to 1996, (Atalay et al., 2011) characterized the buyer–supplier network in the U.S., (Owsiak and Vasquez, 2021) investigates how peaceful country dyads are formed from a territorial perspective, just to name a few. To estimate the influence of geographical proximity on users' probabilities to connect, I implement a series of dyadic Logit regressions to model the data. Following the notation from (De Paula, 2020), a simple dyadic Logit model, for either a directed or undirected network, could be specified as:

$$W_{ij} = 1\left\{D_{ij}\beta_D + X_{ij}^T\beta_X + \epsilon_{ij} \geq 0\right\}$$

where W_{ij} is the connectedness of user i and j, D_{ij} are pair-wise covariates of interests, such as geographical distance between the two users, X_{ij} includes control variables which could be user-specific or dyad-specific variables, and ϵ_{ij} is a logistic random variable. I include a rich set of vertex features as well as a series of constructed dyadic features that reflect homophily in X_{ij}. Summary statistics of the vertex attributes are shown in the following table:

Table 2.1: Summary Statistics of Vertex Attributes

Statistic	Mean	St. Dev.	Min	Max
App Usage in Days	662.70	285.990	4	910
App Version	4.72	0.87	3	7
Platform IOS	0.25	0.43	0	1
Age	30.253	9.95	18.00	99.00
Height(cm)	181.69	153.14	91.00	200.00
Weight(kg)	70.79	14.23	27.00	293.00
Uploaded Photo	0.82	0.39	0.00	1.00
Account Visible	0.97	0.17	0.00	1.00
Has Email	0.70	0.46	0.00	1.00
Photos Rejected by Platform	6.00	26.47	0	1,477
Feed Posts Made in V4	0.63	53.77	0	5,872
Feed Posts Made in V5	7.55	49.70	0	1,996
Feed Posts Made in V6	14.11	70.64	0	2,744
Feed Posts Liked/Commented in V4	44.94	1,101.46	0	84,481
Feed Posts Liked/Commented in V5	789.14	4,130.66	0	116,539
Feed Posts Liked/Commented in V6	1,079.40	4,214.73	0	205,533
Stories Read from Feed in V4	0.05	0.83	0	56
Stories Read from Feed in V5	47.40	228.67	0	10,625
Stories read from Feed in V6	64.41	240.61	0	14,292
Chat Messages Sent in V4	365.88	2,498.43	0	87,302
Chat Messages Sent in V5	5,047.41	11,383.11	0	175,948
Chat Messages Sent in V6	6,181.82	8,414.75	0	116,155
Total Followees in V4	50.51	525.83	0	22,546
Total Followees in V5	842.16	3,080.15	0	167,086
Total Followees in V6	1,166.10	1,656.27	0	49,161
Total Followers	70.21	97.15	0	1,618
Total Followees	70.21	115.53	0	4,243

The sample covers both new users who register an account as late as 4 days ago and old users who have had the account for more than 2 years, including all age groups. Among those who report their races, 47.86% are Asian, 24.40% are White, 7.47% are Mixed, and 20.27% are from other ethnic groups. Self-reported first languages largely coincide with users' countries of origin. 26.76% users speak Turkish, 25.78% speak Thai, 9.77% speak Bahasa Indonesia, 8.83% speak

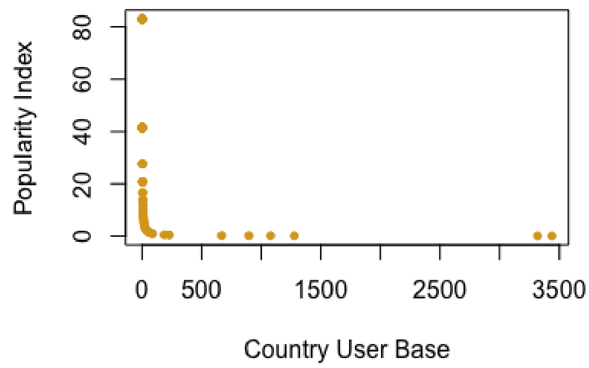


Figure 2.5: Country Popularity

Mandarin, and 8.04% speak English. Users also exhibit different social behaviors. Some lurkers remain inactive on the platform, while others actively send chat messages or interact with a few thousand posts in their feeds. Users' followers range from 0 to 1,618, and the median user has 40 fans.

To investigate the second hypothesis, I define the popularity index for the country as the number of followers from other countries normalized by the number of users from that country, and include it as an additional regressor:

$$Popularity = \frac{Number\ of\ Followers\ Outside\ the\ Country}{Country\ User\ Base}$$

The country popularity index stands for how many times the number of followers outside the country is to the size of the user base or the average number of foreign followers each user from that country has. Country popularity ranges from 0 to over 80, and the smaller the size of the user base, the more popular the country is. I restrict the user base to be larger than 50 and define the 7 countries with the highest popularity index as popular: Germany, France, United States, Mexico, Ukraine, Iran, and Russia.

I apply the model for both the weakly-tied directed network, in which $W_{ij} = 1$ if user i follows j, and the strongly-tied friendship network, in which $W_{ij} = 1$ only if user i and j have followed each other. Apart from the models for all users, to investigate country heterogeneity, I run country-specific models for 6 countries with the largest user base, by restricting the nationality of the follower of each dyad.

Although the dyadic Logit model assumes independence for all dyads and overlooks the strategic part of building connections, it could still replicate stylized attributes of social networks, see (Jochmans, 2018). The challenge that prevents more advanced settings from being adopted is mainly computational. $11,992$ users generate a sample size of more than 143 million dyads, each with more than 100 control variables. The data size makes it impossible to fit into computer memory at one time, and to my best knowledge, there is no off-the-shelf package that runs Logit models at this scale. Thus I used a single-layer neural network, and train it with one smaller batch of the dataset at a time to estimate the parameters.

It is worth mentioning that nationwide travel restrictions and region or city-level stay-at-home orders were enforced for different countries in December 2020. This would rise the probability of people registering and using online social networks, affect the compositions of the user population, or potentially change their social behaviors compared to pre-pandemic times. In addition, 7.14% of users in the network live in Russia and Ukraine. Although the Russo-Ukrainian War has not been escalated until February 2022, the conflict between the two countries, which may include troop incidents, cyber-attacks, and political tensions, could date back to 2014. Online behaviors of users in the area may be affected by following or unfollowing users from other countries.

2.4 Empirical Results

I first present the Logit coefficients for the friendship network, on which only mutual connection between any two users is considered an edge. The model is run for users from the world at first,

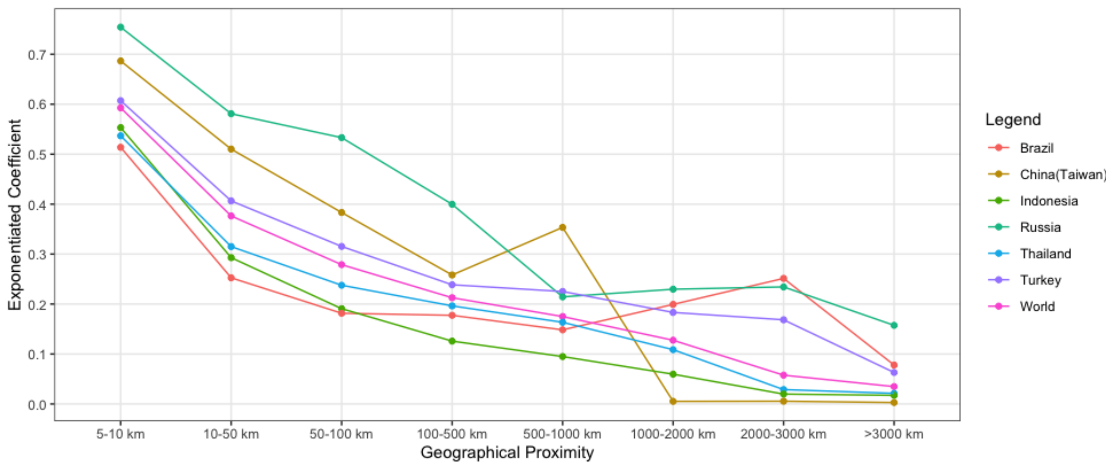


Figure 2.6: Exponentiated Coefficients of Friendship Network by Country

then for users from the 6 largest countries respectively by restricting followers' countries of origin: Thailand, Turkey, Indonesia, China (Taiwan), Brazil, and Russia.

For geographical proximity, the reference level for all models is set to those pairs of users whose distance is within 5 kilometers. Proximity is still of great importance in online social networks, as shown by the significant coefficients, and users are most likely to establish friendships with their neighbors. When the distance between two users becomes further but is still within the practical range to meet offline on a regular basis (100 kilometers), the probability of forming mutual connections declines. Different user behavior patterns could be observed for different countries when the distance between two users goes beyond 100 kilometers. For China (Taiwan), the probability drops until 500 km, then rises a little when the distance is between 500 and 1,000 kilometers. However, this upswing is not necessarily due to changes in user behaviors but rather the lack of users in the sample, given that China (Taiwan) is an island surrounded by the western Pacific Ocean. The connections are mainly built with users from Hong Kong. Taiwanese hardly establish strong ties with users more than 1,000 kilometers away. Another pattern is found for users from the largest country in the world, Russia: users are the most tolerant to longer distances, compared to other countries, within 100 kilometers, but the probability of friendship formation drops significantly if distances get further and catch up with the other countries near 1,000 kilometers. When distance surpasses 1,000 kilometers, the probability of mutually connecting with others grows slowly until the 3,000-kilometer threshold is reached. Besides those who are from another far-off Russian city, plenty of links are formed with users from Turkey. Correspondingly, geographical proximity plays a weaker role for Turkish users who are not from the same place. The probability of making friends remains the same for pairs of users within the range of 100 and 1,000 kilometers and of 1,000 and 3,000 kilometers. The only country in South America, Brazil, shows particular characteristics of how users use the social network app. Within shorter distances, the probability of establishing mutual connections declines sharply in the wake of growing distance, with the largest decreasing amplitude among all countries. When the two users have no chance to meet in person, the probability of following each other does not descend further and even rises again over a long distance. The other two Asian countries, Indonesia and Thailand, behave similarly: friendship gradually and steadily becomes more difficult to build between two people and the rising distance between them.

To transform the scale from log odds to probability, I compute the marginal effects of geographical proximity at mean, for a typical dyad of each country. At the world level, compared with two users who live within 5 kilometers, the probability of the dyad connecting decreases 0.028 when the distance is between 5 and 10 kilometers, and 0.043 when the distance is between 10 and 50 kilometers. Probability reduces along with the increase in distance, and it decays slower for further distances. The table of the marginal effects is shown in B.1.

For covariate effects, users who use the same platform (IOS/Android) have a higher likelihood to connect. This correlation is found to be significant globally and within countries, with the coefficients ranging from 0.11 to 0.30, notably high in Turkey, Brazil, and Russia. This may be due to users who share the same platform having similar online social behaviors or belonging to similar demographic groups.

Residing on the same continent surprisingly reduces the odds of connection among users, except in Brazil and Russia. Looking more closely at national boundaries, being from the same country significantly increases the odds of connection, with particularly strong effects noted in Thailand,

Table 2.2: Logit Coefficient of Countries' Friendship Network

	World	Thailand	Turkey	Taiwan	Indonesia	Brazil	Russia
Same Platform	0.16***	0.11***	0.28***	0.01	0.21***	0.28***	0.30***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Same Continent	-0.73***	-0.70***	-0.99***	-0.68***	-0.70***	1.83***	1.15***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Same Country	1.04***	1.44***	1.41***	0.07***	1.18***	0.03	0.20***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.02)
Same Region	0.34***	0.91***	0.11**	-0.03**	0.06***	-0.08***	0.57***
	(0.01)	(0.02)	(0.05)	(0.01)	(0.01)	(0.02)	(0.07)
Same City	-0.35***	-0.97***	-0.08*	0.47***	0.13***	0.15***	-0.71***
	(0.01)	(0.02)	(0.05)	(0.05)	(0.02)	(0.02)	(0.08)
Age within 5 Years	0.15***	0.15***	0.13***	0.11***	0.14***	0.08***	0.10***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Same Ethnicity	0.35***	0.16***	0.06***	0.20***	0.24***	0.38***	0.17***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Height within 5cm	0.25***	0.28***	0.08***	0.21***	0.10***	0.13***	0.08***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Weight within 5kg	0.10***	0.18***	0.01	0.07**	0.04	0.10**	0.12***
	(0.01)	(0.02)	(0.02)	(0.03)	(0.06)	(0.05)	(0.03)
Same Language	0.31***	0.24***	0.31***	0.10***	0.44***	0.14***	1.05***
	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)	(0.04)	(0.02)

Note: *p<0.1; **p<0.05; ***p<0.01

Turkey, and Indonesia. Region and city-level analysis produce mixed outcomes. While being in the same region generally leads to a higher probability of connection, the effects of being in the same city vary by country. These findings imply a nuanced and potentially complex relationship between geographical proximity and online social networking. The contrasting patterns among various places may reflect diverse social dynamics, such as varying urbanization rates or digital penetration levels.

Apart from geographical homophily factors, physical and cultural attributes also play a role. For instance, age differences of fewer than 5 years, sharing the same ethnicity, and being within 5 centimeters of height all positively influenced the likelihood of connection. The effect was most pronounced when the users speak the same language, particularly in Russia, indicating the vital role of the language barrier in online social interactions.

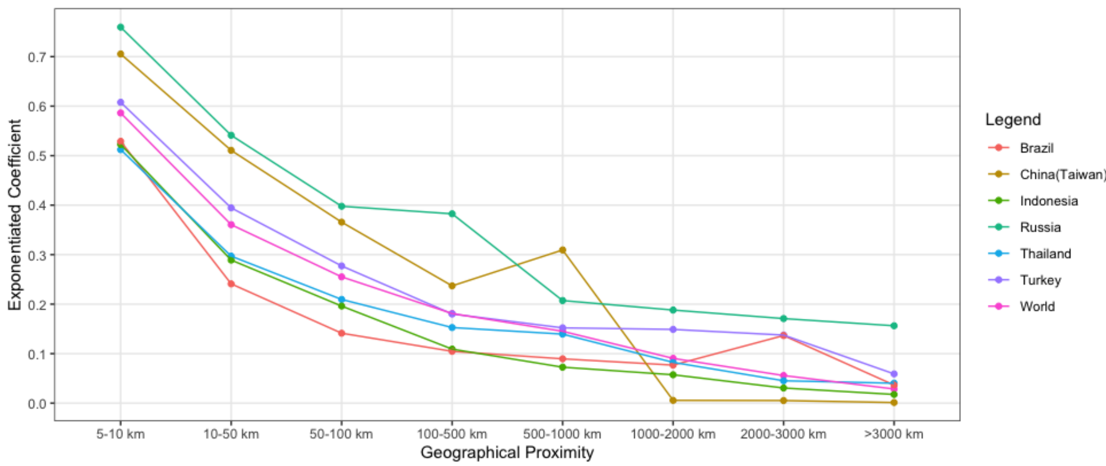


Figure 2.7: Exponentiated Coefficients of Directed Network by Country

When it comes to weaker social ties in a directed follow-unfollow network, country heterogeneity is observed as well. For the three Asian countries, China (Taiwan), Indonesia, and Thailand, probabilities of following another user in response to longer distances change in almost identical fashions with that in a friendship network, albeit with slightly larger distance decays. In Europe, the odds of Russian users following others now also change monotonically with proximity. $1,000$ kilometers is still a threshold near which the probability sharply declines. In South America, users from Brazil still treasure the opportunity to meet up very much whenever possible within a short range. However, when it is infeasible with distances further than 100 kilometers, Brazilians are also interested in forming weaker connections by following users from other places.

Surprisingly, coming from popular countries seems to have no impact or negative impacts on the likelihood of being followed. Users tend to follow others of similar ages, and follower being older tends to increase the likelihood of a follow. Being in the same ethnicity generate higher probability to connect in Thailand, China (Taiwan) and Brazil, but not in other countries. People like to connect with others of similar heights, but the effect is not significant for building connections with users of similar weights. Preference for the height of the followee differs across countries. Taller users, compared to the followers, are more likely to be followed in China (Taiwan), Indonesia, and Brazil, but are less likely to be followed in Thailand and Turkey. User activities in the app, particularly in terms of posts made and stories read, vary widely in their impacts on the connecting likelihood. The coefficients are shown in appendix B.2, B.3, B.4, B.5.

In both the friendship and directed online social networks, users appreciate geographical proximity, and are most likely to connect with others who are within 5 kilometers:

Finding 1. Geographical proximity plays an important role in online social networks, despite the occurrence of country-specific patterns.

Although connections are hard to establish with distance for all countries in general, the decay of odds to form social ties is slower for the friendship network, especially when the distance is beyond 500 kilometers, particularly for Russia and Brazil:

Finding 2. Strong social ties (mutual connections) are less sensitive to the change of physical distance than weak ties (unidirectional connections).

Relative to the unidirectional relationship of a follower and a followee, users who follow each other are tended to send more chat messages, read each other's posts, share common values, and thus yield stronger bounds that help bridge the gap induced by geographical locations.

It is reported that people desire to catch up face-to-face or hang out together from time to time so that the majority of edges cluster within closed areas, e.g., see (Ugander et al., 2011, Ellison, Steinfield and Lampe, 2006). Nevertheless, it is unclear if the intensity of spatial dependence would remain the same at longer distances. In this data, the effects of proximity are lessened for pure net friends who are not able to meet offline, as depicted by the flattening slopes of all curves:

Finding 3. Spatial dependence is stronger when two users may meet in person and becomes looser when distances are longer.

As pointed out in (Graham, 2020), the dyadic Logit model assumes independence across dyads even when they share one or two users in common. It does not usually hold in reality, and uncontrolled interdependence may lead to standard errors and hinder the identification of coefficients. For example, apart from observed covariates, many unobserved user characteristics have not yet been controlled, such as personality, cultural background, etc. It is possible that these factors may also affect link formation, induces correlations among dyads, and thus obscure the identification of coefficients for geographical proximity, e.g., outgoing people may send friend requests to others indifferently regardless of their locations. To allow for correlated dyads and control for unobserved user heterogeneity, (Charbonneau, 2017) and (Graham, 2017) incorporate fixed effects into the model for directed and undirected networks, respectively. Borrowing the idea from panel data literature, the estimators are able to difference-out the unobserved individual fixed effects by focusing on tetrads that satisfy certain conditions.

To illustrate the method, assume user i follows j based on the following rules in a directed network:

$$W_{ij} = 1\left\{D_{ij}\beta_D + X_{ij}^T\beta_X + \alpha_i^{out} + \alpha_j^{in} + \epsilon_{ij} \geq 0\right\}$$

where α_i^{out} and α_j^{in} are unobserved individual fixed effects. To estimate coefficients, consider tetrads composed of four users, i, j, k, and l. The likelihood of l follows j, conditional on either l follows j or l follows k, could be shown that it does not depend on α_k^{out}:

$$P(W_{lj} = 1|D, X, \alpha, W_{lj}+W_{lk} = 1) = \frac{exp\left[(D_{lj} - D_{lk})\beta_D + (X_{lj} - X_{lk})^T\beta_X + \alpha_j^{in} + \alpha_k^{in}\right]}{1 + exp\left[(D_{lj} - D_{lk})\beta_D + (X_{lj} - X_{lk})^T\beta_X + \alpha_j^{in} + \alpha_k^{in}\right]}$$

Similar derivations could be done for events $P(W_{lj} = 1|W_{lj} + W_{lk} = 1)$, $P(W_{lj} = 1|W_{lj} + W_{ij} = 1)$, etc., and the conditional likelihood of l follows j does not depend on any fixed effect:

$$P(W_{lj} = 1|D, X, \alpha, W_{lj} + W_{lk} = 1, W_{ij} + W_{ik} = 1, W_{lj} + W_{ij} = 1) =$$

$$\frac{exp\left[((D_{lj} - D_{lk}) - (D_{ij} - D_{ik}))\beta_D + ((X_{lj} - X_{lk}) - (X_{ij} - X_{ik}))^T\beta_X\right]}{1 + exp\left[((D_{lj} - D_{lk}) - (D_{ij} - D_{ik}))\beta_D + ((X_{lj} - X_{lk}) - (X_{ij} - X_{ik}))^T\beta_X\right]}$$

Although few data points would be used for estimation, the model does not induce the incidental parameter problem and works well in sparse settings. I apply the method to both the directed

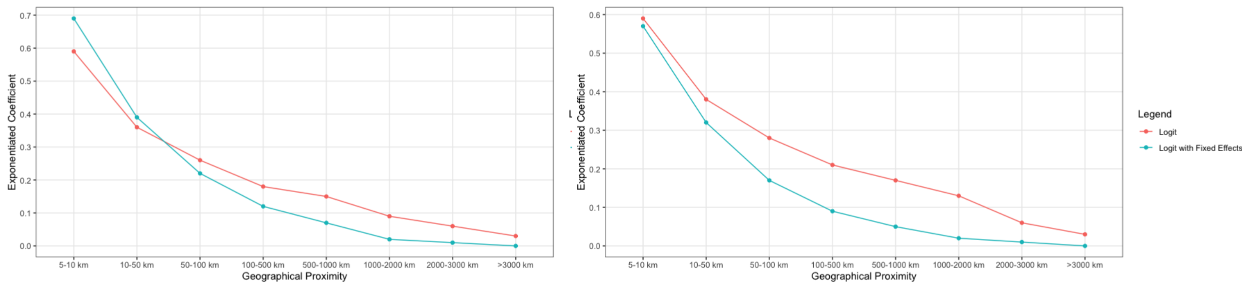


Figure 2.8: Exponentiated Coefficients of Directed (left) and Friendship (right) Network with User Fixed Effects

network and the friendship network, then compare the coefficients with that before adding user fixed effects into the models.

For the friendship network, omitting unobserved characteristics makes the Logit model overstate the odds for all proximity levels. The error is larger for users who are between 50 and 2,000 kilometers, and smaller for users who live in the same place (within 50 kilometers) or beyond 2,000 kilometers. For directed network, the error is generally smaller, and the Logit model would underestimate the effects for users from the same place, but overestimate if the distance between two users is further than 50 kilometers.

2.5 Conclusion

In this paper, I study the role of geographical proximity with respect to link formation in a global online social network with dyadic logit models. I test on three hypotheses: people still prefer to connect with others nearby in a virtual world; users like to connect with those who come from popular countries; users from different countries may exhibit varying behaviors in sending friend requests. Data supports the first and the third hypothesis but has not found significant evidence for the second hypothesis.

Generally, the probabilities to form connections decrease with the rising physical distance between two users, and how they decay depends on users' countries of origin. Proximity is crucial for link formation when the two users are within meeting distance, but becomes weaker if they live in different cities or countries. It also has a lesser effect on a network composed of strong ties, or mutual connections, relative to a network with weak ties, or unidirectional connections.

It is important to note that the potential confounding factor of the app's friend recommendation algorithm, which could also be location-based, could contaminate the empirical results. Isolating the effect of geographical proximity is challenging using the current data, and a few strategies may be adopted when additional information is available. For example, additional control variables for the recommendation algorithm could be included provided the logic of the algorithm is known. Identifying instrumental variables that affect physical distance but are unrelated to the recommendation algorithm is also helpful, e.g. travel time between the two users' locations, policy changes that impact the layout or accessibility of certain areas, etc.

Apart from the potential confounding factors, there is still room left for future work. Firstly, from the graph topology, it can be seen that the network is composed of many regular-type users with fewer connections and a few celebrity users with a great number of followers. There may be reasons to believe that regular type users and celebrity type users behave differently in making connections and thus should be modeled separately. Thirdly, the topology of the network has not been fully taken advantage of in the current model. New variables that reflect both local and global network structure, such as graphon, could be added into the model to capture users' structural tastes for the network, just to name an example. Fourthly, in the present study, all possible dyads are considered, and the assumption is that each user goes over all 11,992 profiles and selects whom to follow. However, users may not be aware of at least some groups of people, e.g. they may only evaluate others who show up in their feeds unless they intentionally search all users worldwide. Thus it is reasonable to develop a model that assumes the users make their follow-unfollow decisions based on some limited consideration sets instead of the entire population. Lastly, it would be interesting to see how social networks in different countries evolve each day, by analyzing more daily snapshots within an observation window and comparing the degree of homophily versus transitivity, as in (Graham, 2016).

Chapter 3

The Impact of COVID-19 on Co-authorship and Economics Scholars' Productivity

The COVID-19 pandemic has disrupted traditional academic collaboration patterns, prompting a unique opportunity to analyze the influence of peer effects and coauthorship dynamics on research output. Using a novel dataset, this paper endeavors to make a first cut at investigating the role of peer effects on the productivity of economics scholars, measured by the number of publications, in both pre-pandemic and pandemic times. Results show that peer effect is significant for the pre-pandemic time but not for the pandemic time. The findings contribute to our understanding of how research collaboration influences knowledge production and may help guide policies aimed at fostering collaboration and enhancing research productivity in the academic community.

3.1 Introduction

Academic productivity is commonly used in academic hiring and promotion decisions, as well as in evaluations of research programs and departments (Gingras, 2016, Bornmann et al., 2008), because a higher number of publications may indicate that the scholar has been active in research and has made significant contributions to the field. In recent years, there has been a growing interest in understanding the factors that contribute to the productivity of scholars in various disciplines. Among these factors, the role of peer effects and co-authorship has emerged as an area of significant importance (Ductor et al., 2014). Scholars who have strong social networks and interactions with colleagues in their field may publish more, because they have access to more resources, including research funding, data, and other research materials. Co-authorship also facilitates knowledge sharing, which could lead to new research ideas and opportunities for publication. For example, (Petersen, 2015) demonstrates that one single strong connection could have a substantial and positive influence on the scholar's productivity and citation rates.

The COVID-19 pandemic has fundamentally altered the way we live, work, and collaborate. As the pandemic forced the closure of universities and research institutions, it inadvertently reshaped the landscape of academic research. Consequently, this unprecedented situation presents a unique opportunity to examine the peer effects and co-authorship dynamics among economics scholars in both pre-pandemic and pandemic times. Working at home may reshape scholars' way of writting, communication, and collaboration. On one hand, scholars have more flexible schedules and greater autonomy over their work. It provides opportunities for virtual collaboration through video conferencing and instant messaging among scholars who locate in different regions or time zones. On the other hand, scholars who work from home may experience greater isolation and have less opportunity for informal interactions and discussions with their co-authors. It may also create distractions and disruptions, such as balancing work responsibilities with home responsibilities.

This paper employs a novel dataset, collected from Google Scholar, that includes scholarly literature and co-authorship networks of economics scholars. I empirically estimate the peer effect of economic scholars on their number of publications during both the pre-pandemic and pandemic periods. The present study is important for several reasons. First, understanding the role of co-authorship in academic productivity provides insights into how research collaborations influence knowledge production and dissemination. Second, by comparing pre-pandemic and pandemic times, this study offers a better understanding of how the COVID-19 crisis has affected academic collaboration patterns and the overall productivity of scholars in the field of economics. Finally, the findings of this study may help guide policies aimed at fostering collaboration and enhancing research productivity in the academic community.

This paper is organized as follows. Section two reviews the relevant literature on peer effects and academic productivity. Section three describes the data and provides summary statistics. Section four presents the model framework and empirical strategy, while Section five discusses the main results. Finally, Section six concludes and offers directions for future research.

3.2 Literature Review

From the methodological perspective, our work is closely related to two main strands of the literature. Firstly, fast-growing literature on peer effects in networks. As reflected by (Manski, 1993), there are two different impacts from peers. One is endogenous peer effects which are the impact of peers' outcomes and the other are contextual peer effects, which are the impact of peers' characteristics. Distinguishing these impacts may be impossible because of the simultaneity in the behavior of interacting agents. (Bramoullé, Djebbari and Fortin, 2009, De Giorgi, Pellizzari and Redaelli, 2010, Lin, 2010, Laschever, 2005) are the initial studies of network interactions. For example, (Bramoullé, Djebbari and Fortin, 2009) builds the benchmark linear-in-means model of peer effects and describes the identification conditions when agents interact through a network assuming peers of peers are not peers. The assumption that the agents' characteristics have an impact on individual outcomes only through their effect on peers' outcomes, provides valid instruments addressing correlated effects. Therefore, endogenous and contextual peer effects are identified. An important insight is that identification depends on the structure of the network itself. (Bramoullé, 2013, Arduini, Patacchini and Rainone, 2019*a*,*b*, Beugnot et al., 2019) investigate heterogeneous peer effects, in which men and women are subject to different peer effects. Individuals could also be subject to different effects from male peers and from female peers, which potentially leads to endogenous peer effects. Concerning heterogeneity of peer effects, (Masten, 2018) incorporates heterogeneity analysis by assuming that endogenous peer effect coefficients are random in a linear-in-means model. He shows these random endogenous peer effects can be point-identified if there is no contextual peer effect for an exogenous characteristic.

The second strand of literature is related to tackling the problem of correlated effects and exploiting the identification possibilities generated by interaction networks. In the literature, researchers have developed at least four broad strategies to address this correlated effects issue: random peers, random shocks, structural endogeneity, and panel data. The first strategy is random peers, who are randomly allocated through natural or designed experiments. For example, (Sacerdote, 2001) looks at Dartmouth College roommates in pairs, triples, or quads among students. (Falk and Ichino, 2006) randomly match workers in pairs in the lab. (De Giorgi, Pellizzari and Redaelli, 2010) look at the choice of a major among Bocconi undergraduates. The insight is with random peers an agent's observed and unobserved characteristics are uncorrelated with their peers' observed and unobserved characteristics. The second strategy is random shocks. For instance, (Dieye, Djebbari and Barrera-Osorio, 2014) uses exogenous variations to study treatment randomization which allows researchers to identify the causal impacts of the treatment and peers' treatments and peers' outcomes even when the network is endogenous within a linear-in-means framework. (Miguel and Kremer, 2004, Kremer and Miguel, 2007, Crépon et al., 2013) study spillover effects. As an individual's potential outcome may depend on the full vector of potential treatments, the causal impact of a randomized treatment cannot be estimated by simply computing the difference in average outcome among treated and untreated individuals. (Hudgens and Halloran, 2008, Vazquez-Bare, 2022) identify spillovers by assuming agents are organized in groups and that spillovers take place within, not between groups, then comparing the outcomes of untreated individuals in treated and untreated groups. The third strategy is called the structural framework. (Goldsmith-Pinkham and Imbens, 2013) first proposed the structural approach to address correlated effects issues. This approach provides a potentially powerful way to control for network endogeneity in peer effect regressions, reminiscent of Heckman's correction for sample network formation simultaneously may allow researchers to recover information on common unobservables. The fourth strategy on peer effects in networks applies to panel data. However, this literature is scarce. A few papers analyze peer effects utilizing panel data, see (Patnam and Sarkar, 2011, Comola and Prina, 2021, De Giorgi, Frederiksen and Pistaferri, 2020). These studies have introduced individual fixed effects but do not address contextual peer effects associated with time-invariant characteristics. This could be viewed as a potentially important limitation of these frameworks and thus is an important future research question.

The model that is employed in the present study belongs to the structural endogeneity framework. It relates to the models that use static games with incomplete information in which, agents act non-cooperatively, see (Harsanyi, 1967, Osborne and Rubinstein, 1994). The assumption of incomplete information of the peer effect models for discrete outcomes is broadly studied, for instance, (Brock and Durlauf, 2001, Bajari et al., 2010, Yang and Lee, 2017, De Paula, 2017). In the literature, agent i's decision is influenced by their own observable characteristics, unobservable individual type, and other agents' choice. The recent study from (Boucher and Houndetoungan, 2020, Houndetoungan, 2022) propose methods to estimate the network's probability distribution using cross-sectional data when the network is imperfectly observed. They construct a network game, and each agent chooses an integer outcome to maximize his or her preference, which contains observed characteristics of the agent and the peers, the difference between the choice of the agent and the peers, a cost function, and a private signal. They prove that under a few assumptions, there is a unique Bayesian Nash Equilibrium for this game, and an estimator is proposed based on pseudo-likelihood maximization.

The literature of empirical papers exploring the relationship between co-authorship and academic productivity has expanded recently. Nonetheless, consensus remains elusive regarding whether this relationship is positive, negative, or insignificant. For example, (Cainelli et al., 2015) demonstrates that economists who are more collaborative are also more productive. Factors such as tenure, age, and geographical variables do not have a significant impact on productivity. (Ductor, 2015) also finds a positive correlation between intellectual collaboration and individual performance, after accounting for endogenous network formation, unobservable heterogeneity, and factors that vary over time. As indirect evidence, (Bosquet and Combes, 2013) identifies that, at the individual level, the average publication quality rises with the average number of authors per paper, individual field diversity, the total number of published papers, and the presence of foreign co-authors. Female and older academics tend to publish less frequently. Conversely, some scholars argue that the relationship between co-authorship and productivity exhibits a negative correlation or lacks statistical significance. (Hollis, 2001) discovers that for a specific scholar, increased co-authorship leads to higher quality, longer, and more frequent publications. However, after adjusting for the number of authors, the relationship between co-authorship and a scholar's attributable output becomes negative. Other research contends that after controlling for article length, journal and author quality, and subject area, scholar fixed effects, the productivity of prior collaborators is not a significant determinant of a researcher's own productivity (Cheng, 2022), or higher quality research (Medoff, 2003). (Oettl, 2012) finds that coauthors of highly helpful scientists that die experience a decrease in output quality but not output quantity.

3.3 Economic Scholars Data

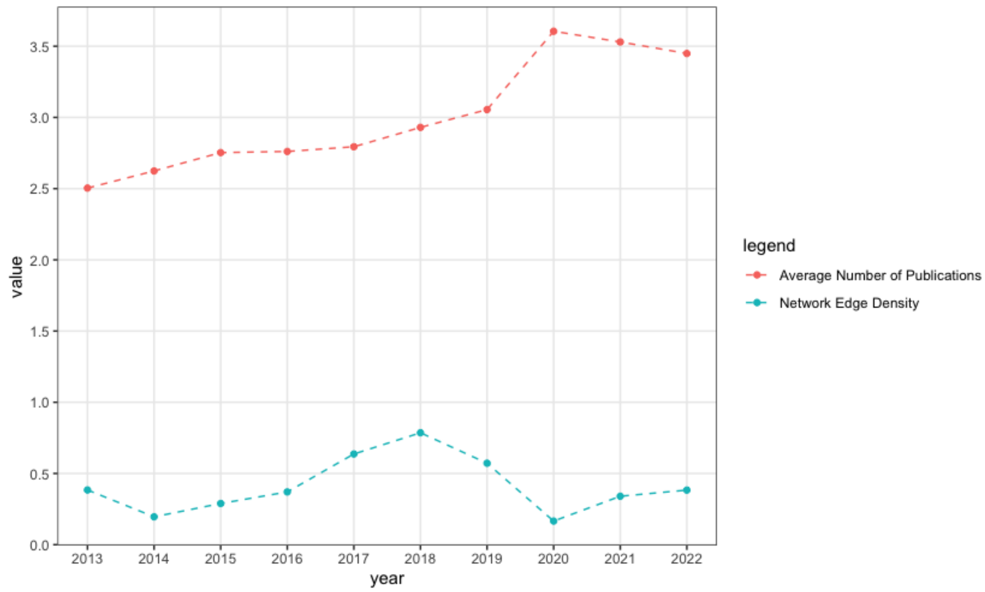
The Economic scholars dataset consists of 1,671 core faculties from the best 50 Economics Schools in the United States based on US News in 2022, who have registered themselves a homepage on Google Scholar. The dataset does not include visiting professors, teaching professors, or lecturers. The homepage provides rich information on the individual's influences and research journey in academia. Besides the names and affiliations, scholars may list their research interests and sub-fields at the head of the page. On the right-hand side, the scholar's cumulative citations, H-index, and I10 index are shown, along with regular coauthors and the histogram of the number of citations and publications each year for the last 10 years. On the left-hand side, a comprehensive list of academic output of the scholar's could be viewed, including the paper's title, journal, authors, year, and the number of citations.

A key feature of the data collected from Google Scholar is its comprehensiveness and strong timeliness. The website aggregates information from a wide range of sources, including journal articles, conference papers, theses and dissertations, technical reports, books and book chapters, patents, working papers, and repositories. Thus compared to using data sources from academic journals, the lagged effect of co-authorship on scholars' academic output due to journals' reviewing process could be largely alleviated.

Based on scholars' publications, I construct an economics scholar co-authorship network each year for the last 10 years. Each node represents a faculty member in the economics department, and there is an edge between any two nodes if they have co-authored at least one paper in a given year. The edge density of a temporal network indicates the prevalence of scholar collaborations in that year, or the percentage of observed co-authorships over all possible collaborations between any two scholars. Economic scholars' productivity, revealed by the annual average numbers of publications, gradually increase since 2013, experience a surge and reaches the top in 2020, then falls back slowly in 2021 and 2022. Network edge density reaches its highest level in 2018, then declines and reaches the bottom in 2020, before it bounces back in 2021 and 2022. Interestingly, scholars' average productivity moves in the same direction as the prevalence of collaborations before 2018, but in the opposite direction from 2019 to 2021.

The experience of economic faculties in academia is 25 years on average as of 2023, which is measured by the difference between 2022 and the year of the scholar's first published article on Google Scholar. African American scholars, identified either directly by their listed nationality in CV or indirectly by the combination of the predicted country of origin using names and faculty pictures, account for 0.9% of the economics faculty. 18% of the faculty in the sample are female. I also label the expertise of economic scholars with their listed sub-fields on the Google homepage. For those who have left the section blank, I fill in the missing values by the fields that are shown on

Figure 3.1: Average Number of Publications and Edge Density by Year



the department faculty website and the scholar's curriculum vitae. The sub-fields are not mutually exclusive, and it is reasonable for scholars to have two, three, or more specialties. For example, a macroeconomist may specify econometrics, monetary policy, international economics, or finance as the expertise as well. In our sample, economic/econometrics theory, macroeconomics, and labor economics are the three top fields that scholars may work on, followed by econometrics, industrial organization, and development economics.

3.4 Network Game with Peer Effects and Incomplete Information

Following the notation in (Houndetoungan, 2022), suppose there is a population of n agents that interact through a network matrix $\mathbf{G}$ with zero diagonal and non-negative elements g_{ij} that represents the proximity of i and j. Each agent i chooses an integer outcome, y_i, to maximize his or her individual utility:

Table 3.1: Summary Statistics of Economics Scholars

Statistic	Mean	St. Dev.	Min	Median	Max
Years in Academia	24.96	15.97	1	22	72
African American	0.009	0.09	0	0	1
Female	0.18	0.39	0	0	1
Number of Publications in 2013	2.504	3.154	0	2	25
Number of Publications in 2014	2.624	3.195	0	2	30
Number of Publications in 2015	2.75	3.38	0	2	27
Number of Publications in 2016	2.76	3.31	0	2	21
Number of Publications in 2017	2.79	3.45	0	2	40
Number of Publications in 2018	2.93	3.43	0	2	40
Number of Publications in 2019	3.06	3.84	0	2	33
Number of Publications in 2020	3.61	4.93	0	2	68
Number of Publications in 2021	3.53	4.46	0	2	55

Table 3.2: Sub-fields of Economics Scholars

	Economics Sub-field (%)
Theory	18.25
Macroeconomics	19.80
Labor Economics	18.01
Econometrics	15.92
Industrial Organization	11.31
Development Economics	11.61
Health Economics	7.60
Financial Economics	10.29

$$U_i\left(y_i, \mathbf{y}_{-i}\right) = \psi_i y_i - c\left(y_i\right) - \frac{\lambda}{2}\left(y_i - \bar{y}_i\right)^2 + e_i\left(y_i\right)$$

In the equation, $\mathbf{y}_{-i} = (y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n)$. Observable characteristics of i and his or her peers is contained in $\psi_i = \mathbf{x}_i{}'\beta + \bar{\mathbf{x}}_\mathbf{i}{}'\gamma$, where $\mathbf{x}_i$ is the vector of the agent's observable characteristics and $\bar{\mathbf{x}}_\mathbf{i}$ is the average characteristics of the peers. Own effects β and contextual effects γ are parameters to be estimated. Function $c\left(\cdot\right)$ captures the cost of choosing y_i, and $\bar{y}_i = \sum_{j=1}^{n} g_{ij} y_j$. The peer effect, $\lambda \geq 0$, is designed to show conformity so that $\frac{\lambda}{2}\left(y_i - \bar{y}_i\right)^2$ represents the social cost that is greater if the difference between the choice of agent i and his or her peers is larger. $(e_i(r))_{r\in\mathbb{N}}$ is a random variable sequence that indicates the agent's private type. The value is observable to i for any $r \in \mathbb{N}$, but not to other agents. Since the types and thus the choices of other agents are not observable, agent i would maximize the expectation of his or her preferences conditional on the information set $I_i = \{\psi_i, \psi_{-i}, \mathbf{g}_i, \mathbf{G}_{-i}\}$:

$$U_i^e(y_i) = \psi_i y_i - c(y_i) - \frac{\lambda}{2}\mathbb{E}_{\bar{y}_i|I_i}\left[(y_i - \bar{y}_i)^2\right] + e_i(y_i)$$

Define Δ as the first difference operator. It is proved in (Houndetoungan, 2022) that under the following three assumptions, there is a unique integer r_0 that maximizes the preference $U_i^e(\cdot)$, and $U_i^e(r) \geq \max\{U_i^e(r-1), U_i^e(r+1)\}$ if and only if $r = r_0$:

Assumption 1. $c(\cdot)$ *is a strictly convex and increasing function.*

Assumption 2. *For any* $r \in \mathbb{N}$, $e_i(r) = e_i(r-1) + \epsilon_i$, *where* $\epsilon_i|I_i$ *are independent and identically follow a continuous symmetric distribution with cdf function* $F_{\epsilon|I}$, *and pdf function* $f_{\epsilon|I}$.

Assumption 3. $\lim\limits_{r\to\infty} r^{-\rho}(\Delta c(r+1) - \Delta c(r)) > 0$, *and* $f_{\epsilon|I}(x) = o(|x|^{-\kappa})$ *at* ∞, *where* $\rho \geq 0$, $(1+\rho)(\kappa-1) > 2$.

The first assumption implies that $\Delta c(r+1) - \Delta c(r) > 0$. The expected payoff is strictly concave and has a global maximum that could be reached at a single point. The second assumption suggests that agents consider the same information ϵ_i for any additional r so that $\Delta e_i(r)$ does not depend on y_i. The third assumption suggests that when y_i is sufficiently high, the cost increases at a minimum rate. The tail of $f_{\epsilon|I}(x)$ needs to decay, and the trade-off condition between ρ and κ guarantees that when $r \to \infty$, the probability of $y_i = r$ converges to 0 at some rate.

Agent i chooses r if and only if $U_i^e(r) \geq U_i^e(r-1)$ and $U_i^e(r) \geq U_i^e(r+1)$. Substituting $U_i^e(\cdot)$ and $e_i(\cdot)$ into the two conditions, we have $-\psi_i - \lambda\bar{y}_i{}^e + a_r \leq \epsilon_i \leq -\psi_i - \lambda\bar{y}_i{}^e + a_{r+1}$, where $a_r = \Delta c(r) + \lambda r - \frac{\lambda}{2}$, $\bar{y}_i{}^e = \sum_{j=1}^n g_{ij} y_j^e$, y_i^e is agent i's rational expected choice conditional on information set I_i. The probability of agent i choosing r, p_{ir}, could readily be written as:

$$p_{ir} = F_{\epsilon|I}\left(\lambda \bar{y}_i{}^e + \psi_i - a_r\right) - F_{\epsilon|I}\left(\lambda \bar{y}_i{}^e + \psi_i - a_{r+1}\right)$$

The expected outcome associate with the belief system $\mathbf{p} = (p_{ir})$ could be written as $y_i^e = \sum_{r=1}^{\infty} r p_{ir} = \sum_{r=1}^{\infty} F_{\epsilon|I}\left(\lambda \bar{y}_i{}^e + \psi_i - a_r\right)$. Although the expected payoff has a global maximum, it is possible that there are multiple expected outcomes and belief systems $\mathbf{p}$. To avoid the multiple rational expected equilibria issue, a threshold for the peer effect needs to be imposed:

Assumption 4. $\lambda < B_c/||\mathbf{G}||_{-\infty}$, *where* $B_c = \left(\max_{u\in\mathbb{R}} \sum_{r=1}^{\infty} f_{\epsilon|I}\left(u - a_r\right)\right)^{-1}$

With the above four assumptions, this game is proved to have a unique Bayesian Nash Equilibrium given by $\mathbf{y}^* = (y_1^*, \cdots, y_n^*)'$, where y_i^* is the maximizer of the expected payoff $U_i^e(\cdot)$.

In the payoff function $\psi_i = \mathbf{x}_i{}'\beta + \bar{\mathbf{x}}_\mathbf{i}{}'\gamma$, let the observable characteristics of agent i and the peers, $\mathbf{x}_i$ and $\bar{\mathbf{x}}_\mathbf{i}$, be $1 \times K$ vectors. Specify $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]'$ as a $n \times K$ matrix, and $\psi = \mathbf{Z\Gamma}$ where $\mathbf{Z} = [\mathbf{X}\ \mathbf{GX}]$ and $\mathbf{\Gamma} = (\beta', \gamma')'$. Define $\delta_r = a_r - a_{r-1}$ for $r \geq 2$ and $\delta_1 = 0$. As $a_r = \Delta c(r) + \lambda r - \frac{\lambda}{2}$, $\delta_r = \Delta\Delta c(r) + \lambda$ for $r \geq 2$. Since $c(\cdot)$ is non-parametric, an infinite number of δ_r needs to be estimated. For identification purposes, the paper assumes the limitation in **Assumption 3** is reached for large r:

Assumption 5. *There exists a positive constant* R, *such that* $\forall r > R$, $\delta_r = (r-1)^{\rho}\bar{\delta} + \lambda$, *where* $\bar{\delta} > 0$, $\rho > 0$.

The probability of agent i choosing r could then be re-written as:

$$p_{ir} = F_{\epsilon|I}\left(\lambda \bar{y}_i{}^e + \mathbf{z}_i'\mathbf{\Gamma} - a_r\right) - F_{\epsilon|I}\left(\lambda \bar{y}_i{}^e + \mathbf{z}_i'\mathbf{\Gamma} - a_{r+1}\right)$$

where $\mathbf{z}_i'$ is $\mathbf{Z}$'s i-th row, $a_0 = -\infty$, $a_r = a_1 + \sum_{k=1}^{r} \delta_k$ for all $r \geq 1$, $\delta_1 = 0$, $\delta_r = (r-1)^{\rho}\bar{\delta} + \lambda$ for all $r > R$. Let $\bar{R}$ be the smallest R for which **Assumption 5** holds. It is shown that under a few additional assumptions in 7, if F_ϵ is known, then $\lambda, \mathbf{\Gamma}, \delta, \bar{\delta}, \rho$ are point identified.

Parameter estimation proceeds with a likelihood approach. Assume ϵ_i follows a standard normal distribution, then the probability p_{ir} would be:

$$p_{ir} = \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \mathbf{z}_i' \mathbf{\Gamma} - a_r\right) - \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \mathbf{z}_i' \mathbf{\Gamma} - a_{r+1}\right)$$

where $\Phi(\cdot)$ is the CDF of standard normal distribution, $\mathbf{L}$ is a mapping, $\mathbf{y}^e = \mathbf{L}(\theta, \mathbf{y}^e) = (l_1(\theta, \mathbf{y}^e), \cdots, l_n(\theta, \mathbf{y}^e))'$, and $l_i(\theta, \mathbf{y}^e) = \sum_{r=1}^{\infty} \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \mathbf{z}_i' \mathbf{\Gamma} - a_r\right)$. For any fixed $\bar{R}$, since $\mathbf{y}^e$ is not observed, it needs to be computed for every value of θ. Alternatively, the parameters could be estimated using the NPL algorithm in (Aguirregabiria and Mira, 2007), which takes advantage of an iterative process. The algorithm maximizes a pseudo-likelihood function:

$$L_n\left(\theta, \mathbf{y}^e\right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{r=0}^{\infty} d_{ir} log(p_{ir})$$

where $\theta = \left(log(\lambda), \mathbf{\Gamma}', log(\tilde{\delta}'), log(\bar{\delta}), log(\rho)\right)'$, $d_{ir} = 1$ if $y_i = r$ and 0 otherwise. It starts by guessing a set of initial probabilities for each agent's choices, and then updating these probabilities in each iteration until the parameters and probabilities converge to a stable solution, e.g. $||\theta_{(t)} - \theta_{(t-1)}||_1$ and $||\mathbf{y}_{(t)} - \mathbf{y}_{(t-1)}||_1$ are less than 10^{-4}.

The model also takes into account the endogeneity problem induced by agents' unobserved characteristics. For instance, in our application, scholars' familiarity with coding, or their extent to communicate with others could affect both which scholars they may collaborate with ($\mathbf{G}$) and their own number of publications ($\mathbf{y}$). Let the latent utility of scholar i and j being coauthors be $\mathbf{g}_{ij}^* = \ddot{\mathbf{x}}_{ij}\bar{\beta} + \mu_i + \nu_j + \eta_{ij}$, where $\ddot{\mathbf{x}}_{ij}$ contains dyadic variables, mu_i and ν_j are individual fixed effects. The probability of scholar i and j being coauthors could be modelled as:

$$P_{ij} = \mathbb{P}(\mathbf{g}_{ij}^* > 0) = F_\eta\left(\ddot{\mathbf{x}}_{ij}\bar{\beta} + \mu_i + \nu_j\right)$$

The fixed effects are assumed to be unobservable for the researcher, but observable for the scholars so that they are included in the information set.

Assumption 6. *For continuous function h_ϵ, $\epsilon_i = h_\epsilon(\mu_i, \nu_i, \bar{\mu}_i, \bar{\nu}_i) + \epsilon_i^*$, where ϵ_i^* is independent of* $\mathbf{Z}$ *and* $\mathbf{G}$, $\bar{\mu}_i = \sum_{j=1}^n \mathbf{g}_{ij}\mu_j$, $\bar{\nu}_i = \sum_{j=1}^n \mathbf{g}_{ij}\nu_j$.

The ϵ could be replaced by $h_\epsilon(\mu_i, \nu_i, \bar{\mu}_i, \bar{\nu}_i) + \epsilon_i^*$. Adapting **Assumption 2** and **Assumption 3** to ϵ_i^*, the defined Bayesian Nash Equilibrium is still valid. The estimator computes $\hat{\mu}_i$ and $\hat{\nu}_i$ using a standard Logit in the first step, then substitute the estimated values for μ_i and nu_i in the second step. Function h_ϵ is approximated using a sieve method. I refer readers that are interested in the model and technical details to the original paper (Houndetoungan, 2022).

The Economics scholar data contains the annual number of publications for each scholar from 2018 to 2021. I split the sample into pre-Covid and Covid periods by time (2018-2019, 2020-2021), and apply the model to each of these 2-year periods to estimate the peer effect. Specifically, I define y_i as the total number of publications of scholar i within each 2-year period. Data is segmented in this way because productivity and collaboration are observed to move in different directions before and after 2019. The pandemic and schools' transition to virtual learning in 2020-2021 also inevitably affect scholars' productivity and the way they collaborate, so the peer effect λ is expected to have some change. I keep the length of each period the same to ensure that the estimated results in the two models are comparable. I include gender, an indicator for African American scholars, expertise, recent productivity indicated by the average number of publications of the scholar in the previous 3 years, total citations up to the first year of each period, and academic experience in the observable characteristics. I discretize scholars' recent productivity, total citations, and academic experience, and set scholars with an average number of 0 or 1 publication in the previous 3 years, less than 100 citations up to the first year of each period, and less than 10 years of experience, as the reference levels respectively. In the presence of productivity rise during Covid times, I construct an additional Covid index for the period 2019-2021 in order to control for the extent to which the scholars' publications after the outbreak of the epidemic are related to the Covid context. I use topic modeling to predict the topic of each paper published within these three years based on the title and abstract. The model is capable of recognizing papers that are inspired by the pandemic from other publications, see the table of predicted topics and the words with the highest conditional probability for each topic in the following table. I define a paper is Covid-related if its probability of belonging to the "Covid-19" topic is higher than 50%. The Covid index for each scholar is calculated as:

$$\text{Covid Index} = \frac{\text{Number of Covid-related Publications in 2019-2021}}{\text{Number of Publications in 2019-2021}}$$

The mean Covid index for economics scholars is 0.11, meaning that on average, 11% of the papers written in 2019-2021 by an economics faculty are inspired by the pandemic.

Table 3.3: Predicted Topics of Economic Publications in 2019-2021

Topic	Key Words	Percentage(%)
Covid-19	Covid-19, Health, Pandemic, Impact, Effects	26.05
Macro/Finance/International	Financial, Trade, Monetary, Policy, Experimental	19.97
Micro/Theory/Metrics/IO	Market, Theory, Labor, Learning, Estimation	27.81
Development/Social Science/Others	Inequality, Gender, Mobility, Work, Review	26.17

Social interaction matrix $\mathbf{G}$ is a row-normalized version of the adjacency matrix $\mathbf{W} = [w_{ij}]$, in which $w_{ij} = 1$ if scholar i and j has collaborated on at least 2 papers within the 3-year period and 0 otherwise. I rule out the case in which two scholars coauthored a single paper to focus on stable relationships. For example, if there are three scholars i, j and l, scholar i has worked with both j and l on more than one paper during the 3-year period. Then $\mathbf{G}$ is specified in the following way:

$$\mathbf{G} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Each agent considers the average characteristics of his or her coauthors. For example, for the payoff function $U_i(y_i)$ in the above example, $\bar{y}_i = \frac{1}{2}y_j + \frac{1}{2}y_l$. Peer effect λ indicates how much "peer pressure" a scholar may have due to the difference between the scholar's own productivity and the average productivity of the co-authors.

To account for the endogeneity of network formation, in the first stage, I estimate scholars' unobserved characteristics using a dyadic Logit model. In the network formation model, from the school level, I control for homophily of the same department, and the same US News Ranking (Top 10, 11-20, 21-30, 31-40, 41-50). From the individual level, I consider the differences between the two scholars' academic experience in years, total citations up to the first year of the period, the average number of publications each year during the previous three years, and the number of total publications. I include dummies to indicate whether the dyad involves at least one female author, and at least one African American scholar. I also control for the two scholars' common research interests, measured by the number of fields that are listed on both scholars' websites. In the second stage, I include individual fixed effects as additional control variables to evaluate the peer effect.

In practice, the value of $\bar{R}$ needs to be specified by the researcher in advance. Following the suggestion of the author, I experiment with the value of $\bar{R}$ by increasing it from 2 until the change of the estimated parameters is not significant, or it reaches $\max(y) - 2$. The estimation is done using the R package provided by the author.

3.5 Empirical Results

I estimate two models for the 2-year pre-pandemic period (2018-2019) and the 2-year in-pandemic period (2020-2021) respectively. For the third model, I include the Covid index as an additional regressor for the in-pandemic model.

The peer effect is significant for pre-Covid times, but not for Covid times. This implies that economic scholars exhibit conformity in the number of publications before the schools switch to virtual mode. However, there is a sign that while working at home, economic scholars are cooperating with a more diverse group of co-authors in terms of productivity, i.e. more collaborations among prolific scholars and scholars who publish less are expected.

Table 3.4: Peer Effect on Scholars' Number of Publications in Non-Covid and Covid Times

	Pre-Covid (2018-2019)	Covid (2020-2021)	Covid + Covid Index (2020-2021)
λ	0.10^{***}	0.01	0.02
	(0.03)	(0.03)	(0.03)
Note:			*p<0.1; **p<0.05; ***p<0.01

Recent productivity, represented by the scholar's average number of publications during the previous three years, is an important predictor of a scholar's productivity in the future. Scholars who are more prolific in the past 3 years would be more productive in the future. Compared to scholars whose total citations are below 2,000, scholars who have more citations are likely to publish more. Scholars' years in academia significantly affect their productivity during Covid times. Emerging scholars produce relatively more than senior scholars. Gender plays a significant role only during the pre-Covid time.

Productivity is also related to the economics sub-fields that scholars work in. For example, econometricians publish relatively fewer papers in 2018-2019. Scholars who work in health economics are more prolific, and the effect is stronger during Covid times. The significant and positive Covid index implies that during 2019-2021, the higher proportion of a scholar's papers is related to Covid, the more publications he or she would have. After controlling for the Covid index, the coefficient of scholars working in the field of health economics drops from 0.33 to 0.30.

Contextual effects are not significant in general, see table in C.1. During Covid times, collaborating with macroeconomists and established scholars who have 100-2,000 citations help publish more papers.

Table 3.5: Own Effects on Scholars' Number of Publications in Non-Covid and Covid Times

	Pre-Covid (2018-2019)	Covid (2020-2021)	Covid + Covid Index (2020-2021)
2-4 Publications Per Year	0.40^{***}	0.28^{***}	0.27^{***}
	(0.07)	(0.06)	(0.06)
5-9 Publications Per Year	1.22^{***}	1.17^{***}	1.17^{***}
	(0.09)	(0.09)	(0.09)
10+ Publications Per Year	2.38^{***}	2.46^{***}	2.46^{***}
	(0.16)	(0.16)	(0.16)
100-499 Citations	-0.15^{*}	-0.21^{**}	-0.21^{**}
	(0.08)	(0.09)	(0.09)
500-1,999 Citations	-0.02	-0.11	-0.11
	(0.07)	(0.07)	(0.07)
2,000-4,999 Citations	0.48^{***}	0.38^{***}	0.37^{***}
	(0.09)	(0.10)	(0.10)
5,000-9,999 Citations	0.71^{***}	0.59^{***}	0.58^{***}
	(0.13)	(0.13)	(0.13)
10,000-19,999 Citations	0.77^{***}	0.86^{***}	0.84^{***}
	(0.15)	(0.14)	(0.14)
20,000+ Citations	1.81^{***}	1.53^{***}	1.52^{***}
	(0.18)	(0.16)	(0.16)
Experience 10-20 Years	-0.01	-0.12	-0.11
	(0.08)	(0.08)	(0.08)
Experience 20-30 Years	-0.18^{*}	-0.28^{***}	-0.27^{***}
	(0.09)	(0.10)	(0.10)
Experience 30-40 Years	-0.12	-0.47^{***}	-0.46^{***}
	(0.10)	(0.11)	(0.11)
Experience 40-50 Years	-0.05	-0.30^{**}	-0.29^{**}
	(0.12)	(0.13)	(0.13)
Experience 50-60 Years	0.25^{*}	-0.36^{**}	-0.35^{**}
	(0.15)	(0.15)	(0.15))
Experience 60+ Years	0.27	-0.39^{**}	-0.39^{**}
	(0.21)	(0.18)	(0.18)
Covid Index			0.27^{*}
			(0.16)
African American	0.33	-0.07	-0.06
	(0.29)	(0.29)	(0.29)
Female	-0.13^{*}	0.04	0.04
	(0.07)	(0.07)	(0.07)
Field: Theory	-0.10	0.09	0.10
	(0.07)	(0.07)	(0.07)
Field: Macro	-0.04	-0.001	0.01
	(0.07)	(0.07)	(0.07)
Field: Labor	-0.09	-0.02	-0.04
	(0.07)	(0.07)	(0.07)
Field: Metrics	-0.22^{***}	0.04	0.05
	(0.07)	(0.07)	(0.07)
Field: Industrial Organization	-0.02	-0.06	-0.06
	(0.08)	(0.08)	(0.08)
Field: Development	-0.04	0.16^{*}	0.16^{*}
	(0.08)	(0.08)	(0.08)
Field: Health	0.21^{**}	0.33^{***}	0.30^{***}
	(0.10)	(0.10)	(0.10)
Field: Finance	-0.13	-0.01	-0.01
	(0.09)	(0.09)	(0.09)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

3.6 Conclusion

In this study, I examine the influence of peer effects on the productivity of economics scholars in both pre-pandemic and pandemic periods. Findings reveal that scholars tend to coauthor with others with similar productivity during the pre-pandemic time, but this conformity is not found during the pandemic time. Productivity during the previous 3 years helps predict productivity in the near future. Citation count is positively correlated with scholars' productivity at both times, while academic experience only affects the pandemic time. Female scholars are reported to publish less during the pre-pandemic time, but the effect vanishes during the pandemic time. This paper contributes to the growing body of literature on peer effects and academic productivity, and the insights help inform policies aimed at fostering research collaboration and enhancing productivity within the academic community.

While our research offers important insights, it is not without limitations. For example, in the present study, I assess an economics scholar's productivity based on the number of publications. This may be questionable because factors like the quality of the research and the type of publications are not considered. More comprehensive measures could be used by including citation counts, journal impact factors, type of publications, altmetrics, etc. Relative to other data sources, although Google Scholar has several advantages, such as broader coverage of different types of publication, the information being up-to-date, etc., it may offer inaccurate author profiles and results of inconsistent accuracy (Falagas et al., 2008). Future research may benefit from incorporating additional data sources from journal websites or other online platforms to ensure a more comprehensive and accurate assessment of scholars' productivity and co-authorship patterns. In addition, I only investigate the peer effect within a four years period. As the pandemic continues to affect the global academic community, it would be crucial to monitor the long-term implications of these changing collaboration patterns and their impact on research productivity. Further exploration of factors that facilitate or hinder research collaboration during such crises can help guide the development of effective strategies and policies to support and strengthen the academic community in times of unprecedented challenges.

Bibliography

Aga, Gemechu, and David Francis. 2017. "As the market churns: productivity and firm exit in developing countries." *Small Business Economics*, 49: 379–403.

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential estimation of dynamic discrete games." *Econometrica*, 75(1): 1–53.

Albert, Réka, Hawoong Jeong, and Albert-László Barabási. 1999. "Diameter of the world-wide web." *nature*, 401(6749): 130–131.

Amemiya, Takeshi. 1974. "The nonlinear two-stage least-squares estimator." *Journal of econometrics*, 2(2): 105–110.

Amemiya, Takeshi. 1975. "The nonlinear limited-information maximum-likelihood estimator and the modified nonlinear two-stage least-squares estimator." *Journal of Econometrics*, 3(4): 375–386.

Anderson, James E. 2011. "The Gravity Model: Annual Review of Economics."

Anselin, Luc. 1988. *Spatial econometrics: methods and models.* Vol. 4, Springer Science & Business Media.

Anselin, Luc, Raymond Florax, and Sergio J Rey. 2013. *Advances in spatial econometrics: methodology, tools and applications.* Springer Science & Business Media.

Arcuri, Giuseppe, and Nadine Levratto. 2020. "Early stage SME bankruptcy: does the local banking market matter?" *Small Business Economics*, 54: 421–436.

Arcuri, Giuseppe, Marc Brunetto, and Nadine Levratto. 2019. "Spatial patterns and determinants of firm exit: an empirical analysis on France." *The Annals of Regional Science*, 62(1): 99–118.

Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone. 2019*a*. "Identification and Estimation of Network Models with Between and Within Groups Interactions."

Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone. 2019*b*. "Treatment effects with heterogeneous externalities." *Journal of Business & Economic Statistics*, 38(4): 826–838.

Atalay, Enghin, Ali Hortacsu, James Roberts, and Chad Syverson. 2011. "Network structure of production." *Proceedings of the National Academy of Sciences*, 108(13): 5199–5202.

Audretsch, David B, and Maryann P Feldman. 1996. "R&D spillovers and the geography of innovation and production." *The American economic review*, 86(3): 630–640.

Audretsch, David B, Dirk Dohse, and Annekatrin Niebuhr. 2015. "Regional unemployment structure and new firm formation." *Papers in Regional Science*, 94: S115–S138.

Backstrom, Lars, Eric Sun, and Cameron Marlow. 2010. "Find me if you can: improving geographical prediction with social and spatial proximity." 61–70.

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov. 2010. "Estimating static models of strategic interactions." *Journal of Business & Economic Statistics*, 28(4): 469–482.

Baldwin, John R, and Paul K Gorecki. 1991. "Firm entry and exit in the Canadian manufacturing sector, 1970-1982." *Canadian Journal of Economics*, 300–323.

Barney, Jay. 1991. "Firm resources and sustained competitive advantage." *Journal of management*, 17(1): 99–120.

Basile, Roberto, Rosanna Pittiglio, and Filippo Reganati. 2017. "Do agglomeration externalities affect firm survival?" *Regional Studies*, 51(4): 548–562.

Baumgardner, James R. 1988. "The division of labor, local markets, and worker organization." *Journal of Political Economy*, 96(3): 509–527.

Benito, Gabriel RG. 1997. "Divestment of foreign production operations." *Applied economics*, 29(10): 1365–1378.

Beron, Kurt J, and Wim PM Vijverberg. 2004. "Probit in a spatial context: a Monte Carlo analysis." In *Advances in spatial econometrics*. 169–195. Springer.

Beugnot, Julie, Bernard Fortin, Guy Lacroix, and Marie Claire Villeval. 2019. "Gender and peer effects on performance in social networks." *European Economic Review*, 113: 207–224.

Blake, Robert R, Clifton C Rhead, Bryant Wedge, and Jane Srygley Mouton. 1956. "Housing architecture and social interaction." *Sociometry*, 19(2): 133–139.

Boase, Jeffrey, John Horrigan, Barry Wellman, and Lee Rainie. 2006. "The strength of internet ties." *Pew Research Center's Internet & American Life Project*.

Bornmann, Lutz, Rüdiger Mutz, Christoph Neuhaus, and Hans-Dieter Daniel. 2008. "Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results." *Ethics in science and environmental politics*, 8(1): 93–102.

Bosquet, Clément, and Pierre-Philippe Combes. 2013. "Do large departments make academics more productive? Agglomeration and peer effects in research."

Boucher, Vincent, and Aristide Houndetoungan. 2020. *Estimating peer effects using partial network data.* Centre de recherche sur les risques les enjeux économiques et les politiques ....

Bown, Chad P. 2019. "US-China trade war: The guns of August." *Trade and investment policy watch*, 20.

Bramoullé, Yann. 2013. "A comment on social networks and the identification of peer effects by P. Goldsmith-Pinkham & GW Imbens." *Journal of Business and Economic Statistics*, 31(3): 264–266.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of econometrics*, 150(1): 41–55.

Brock, William A, and Steven N Durlauf. 2001. "Discrete choice with social interactions." *The Review of Economic Studies*, 68(2): 235–260.

Butts, Carter T, Ryan M Acton, John R Hipp, and Nicholas N Nagle. 2012. "Geographical variability and network structure." *Social Networks*, 34(1): 82–100.

Cainelli, Giulio, Mario A Maggioni, T Erika Uberti, and Annunziata De Felice. 2015. "The strength of strong ties: How co-authorship affect productivity of academic economists?" *Scientometrics*, 102: 673–699.

Cainelli, Giulio, Sandro Montresor, and Giuseppe Vittucci Marzetti. 2014. "Spatial agglomeration and firm exit: a spatial dynamic analysis for Italian provinces." *Small Business Economics*, 43(1): 213–228.

Cairncross, Frances. 2001. *The death of distance: How the communications revolution is changing our lives.*

Calabrese, Raffaella, and Johan A Elkink. 2014. "Estimators of binary spatial autoregressive models: A Monte Carlo study." *Journal of Regional Science*, 54(4): 664–687.

Calá, Carla Daniela, Miguel Manjón-Antolín, and Josep-Maria Arauzo-Carod. 2016. "Regional determinants of firm entry in a developing country." *Papers in regional science*, 95(2): 259–279.

Cardon, Peter W, Bryan Marshall, Jeongil Choi, Maha M El-Shinnaway, Matthew North, Lars Svensson, Sujie Wang, Daniel T Norris, Lixin Cui, Natalya Goreva, et al. 2009. "Online and offline social ties of social network website users: An exploratory study in eleven societies." *Journal of Computer Information Systems*, 50(1): 54–64.

Cefis, Elena, Cristina Bettinelli, Alex Coad, and Orietta Marsili. 2022. "Understanding firm exit: a systematic literature review." *Small Business Economics*, 59(2): 423–446.

Charbonneau, Karyne B. 2017. "Multiple fixed effects in binary response panel data models." *The Econometrics Journal*, 20(3): S1–S13.

Cheng, Wei. 2022. “Productivity spillovers in endogenous coauthor networks.” *Empirical Economics*, 63(6): 3159–3183.

Chen, Wenhong, Jeffrey Boase, and Barry Wellman. 2002. “The global villagers.” *The Internet in everyday life*, 74–113.

Cheratian, Iman, Saleh Goltabar, and Carla Daniela Calá. 2021. “Spatial drivers of firm entry in Iran.” *The Annals of Regional Science*, 66: 463–496.

Chin, Alvin, Bin Xu, and Hao Wang. 2013. “Who should I add as a" friend"? A study of friend recommendations using proximity and homophily.” 1–7.

Chin, Alvin, Bin Xu, Hao Wang, and Xia Wang. 2012. “Linking people through physical proximity in a conference.” 13–20.

Comola, Margherita, and Silvia Prina. 2021. “Treatment effect accounting for network changes.” *The Review of Economics and Statistics*, 103(3): 597–604.

Cotei, Carmen, and Joseph Farhat. 2018. “The M&A exit outcomes of new, young firms.” *Small Business Economics*, 50: 545–567.

Coucke, Kristien, and Leo Sleuwaegen. 2008. “Offshoring as a survival strategy: evidence from manufacturing firms in Belgium.” *Journal of International Business Studies*, 39: 1261–1277.

Craioveanu, Mihaela, and Dek Terrell. 2016. “The impact of storms on firm survival: a Bayesian spatial econometric model for firm survival.” In *Spatial Econometrics: Qualitative and Limited Dependent Variables*. Vol. 37, 81–118. Emerald Group Publishing Limited.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. “Do labor market policies have displacement effects? Evidence from a clustered randomized experiment.” *The quarterly journal of economics*, 128(2): 531–580.

De Giorgi, Giacomo, Anders Frederiksen, and Luigi Pistaferri. 2020. “Consumption network effects.” *The Review of Economic Studies*, 87(1): 130–163.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. “Identification of social interactions through partially overlapping peer groups.” *American Economic Journal: Applied Economics*, 2(2): 241–275.

De Paula, Aureo. 2017. “Econometrics of network models.” 268–323, Cambridge University Press Cambridge.

De Paula, Áureo. 2020. “Econometric models of network formation.” *Annual Review of Economics*, 12: 775–799.

Dieye, Rokhaya, Habiba Djebbari, and Felipe Barrera-Osorio. 2014. “Accounting for peer effects in treatment response.”

Ductor, Lorenzo. 2015. "Does co-authorship lead to higher academic productivity?" *Oxford Bulletin of Economics and Statistics*, 77(3): 385–407.

Ductor, Lorenzo, Marcel Fafchamps, Sanjeev Goyal, and Marco J Van der Leij. 2014. "Social networks and research output." *Review of Economics and Statistics*, 96(5): 936–948.

Dunning, John H. 1980. "Toward an eclectic theory of international production: Some empirical tests." *Journal of international business studies*, 11: 9–31.

Duranton, Gilles, and Diego Puga. 2001. "Nursery cities: urban diversity, process innovation, and the life cycle of products." *American Economic Review*, 91(5): 1454–1477.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of urban agglomeration economies." In *Handbook of regional and urban economics*. Vol. 4, 2063–2117. Elsevier.

Ellison, Nicole, Charles Steinfield, and Cliff Lampe. 2006. "Spatially bounded online social networks and social capital." *International Communication Association*, 36(1-37).

Fafchamps, Marcel, and Matthias Schündeln. 2013. "Local financial development and firm performance: Evidence from Morocco." *Journal of Development Economics*, 103: 15–28.

Falagas, Matthew E, Eleni I Pitsouni, George A Malietzis, and Georgios Pappas. 2008. "Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses." *The FASEB journal*, 22(2): 338–342.

Falk, Armin, and Andrea Ichino. 2006. "Clean evidence on peer effects." *Journal of labor economics*, 24(1): 39–57.

Faloutsos, Michalis, Petros Faloutsos, and Christos Faloutsos. 1999. "On power-law relationships of the internet topology." *ACM SIGCOMM computer communication review*, 29(4): 251–262.

Fan, C Cindy, and Allen J Scott. 2003. "Industrial agglomeration and development: a survey of spatial economic issues in East Asia and a statistical analysis of Chinese regions." *Economic geography*, 79(3): 295–319.

Ferragina, Anna M, and Fernanda Mazzotta. 2015. "Agglomeration economies in Italy: impact on heterogeneous firms' exit in a multilevel framework." *Economia e Politica Industriale*, 42: 395–440.

Fleming, Mark M. 2004. "Techniques for estimating spatially dependent discrete choice models." In *Advances in spatial econometrics*. 145–168. Springer.

Gingras, Yves. 2016. *Bibliometrics and research evaluation: Uses and abuses.* Mit Press.

Goktan, M Sinan, Robert Kieschnick, and Rabih Moussawi. 2018. "Corporate governance and firm survival." *Financial Review*, 53(2): 209–253.

Goldenberg, Jacob, and Moshe Levy. 2009. "Distance is not dead: Social interaction and geographical distance in the internet era." *arXiv preprint arXiv:0906.3202*.

Goldsmith-Pinkham, Paul, and Guido W Imbens. 2013. "Social networks and the identification of peer effects." *Journal of Business & Economic Statistics*, 31(3): 253–264.

Graham, Bryan S. 2016. "Homophily and transitivity in dynamic network formation." National Bureau of Economic Research.

Graham, Bryan S. 2017. "An econometric model of network formation with degree heterogeneity." *Econometrica*, 85(4): 1033–1063.

Graham, Bryan S. 2020. "Dyadic regression." *The Econometric Analysis of Network Data*, 23–40.

Harsanyi, John C. 1967. "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model." *Management science*, 14(3): 159–182.

Hipp, John R, and Andrew J Perrin. 2009. "The simultaneous effect of social distance and physical distance on the formation of neighborhood ties." *City & Community*, 8(1): 5–25.

Hollis, Aidan. 2001. "Co-authorship and the output of academic economists." *Labour economics*, 8(4): 503–530.

Houndetoungan, Elysée Aristide. 2022. "Count data models with social interactions under rational expectations." *Unpublished manuscript, Thema, Cergy Paris University*.

Hudgens, Michael G, and M Elizabeth Halloran. 2008. "Toward causal inference with interference." *Journal of the American Statistical Association*, 103(482): 832–842.

Huiban, Jean-Pierre. 2011. "The spatial demography of new plants: urban creation and rural survival." *Small Business Economics*, 37(1): 73–86.

Jochmans, Koen. 2018. "Semiparametric analysis of network formation." *Journal of Business & Economic Statistics*, 36(4): 705–713.

Jovanovic, Boyan, and Rafael Rob. 1989. "The growth and diffusion of knowledge." *The Review of Economic Studies*, 56(4): 569–582.

Kim, Yoojung, Dongyoung Sohn, and Sejung Marina Choi. 2011. "Cultural difference in motivations for using social network sites: A comparative study of American and Korean college students." *Computers in human behavior*, 27(1): 365–372.

Klier, Thomas, and Daniel P McMillen. 2008. "Clustering of auto supplier plants in the United States: generalized method of moments spatial logit for large samples." *Journal of Business & Economic Statistics*, 26(4): 460–471.

Kremer, Michael, and Edward Miguel. 2007. "The illusion of sustainability." *The Quarterly journal of economics*, 122(3): 1007–1065.

Laschever, Ron. 2005. “The doughboys network: social interactions and labor market outcomes of World War I veterans.” *Unpublished manuscript, Northwestern University*, 48.

Latané, Bibb, James H Liu, Andrzej Nowak, Michael Bonevento, and Long Zheng. 1995. “Distance matters: Physical space and social impact.” *Personality and Social Psychology Bulletin*, 21(8): 795–805.

Lee, Donghun, and Ravi Madhavan. 2010. “Divestiture and firm performance: A meta-analysis.” *Journal of Management*, 36(6): 1345–1371.

Lee, In Hyeock, Eunsuk Hong, and Laixiang Sun. 2013. “Regional knowledge production and entrepreneurial firm creation: Spatial Dynamic Analyses.” *Journal of Business Research*, 66(10): 2106–2115.

LeSage, James, and Robert Kelley Pace. 2009. *Introduction to spatial econometrics.* Chapman and Hall/CRC.

LeSage, James P. 2000. “Bayesian estimation of limited dependent variable spatial autoregressive models.” *Geographical Analysis*, 32(1): 19–35.

Levratto, Nadine, and Aziza Garsaa. 2016. “Does the employment growth rate depend on the local context? An analysis of French industrial establishments over the 2004-2010 period.” *Revue d'économie industrielle*, , (153): 47–89.

Levratto, Nadine, and Denis Carré. 2014. “Does firm creation depend on local context? a focus on the neighbouring effects.” *Réseau de Recherchesurl'Innovation/Research Network on Innovation 2014*.

Liben-Nowell, David, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2005. “Geographic routing in social networks.” *Proceedings of the National Academy of Sciences*, 102(33): 11623–11628.

Lin, Xu. 2010. “Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables.” *Journal of Labor Economics*, 28(4): 825–860.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem.” *The Review of Economic Studies*, 60(3): 531–542.

Marshall, Alfred. 2009. *Principles of economics: unabridged eighth edition.* Cosimo, Inc.

Masten, Matthew A. 2018. “Random coefficients on endogenous variables in simultaneous equations models.” *The Review of Economic Studies*, 85(2): 1193–1250.

Mata, José, and Ernesto Freitas. 2012. “Foreignness and exit over the life cycle of firms.” *Journal of International Business Studies*, 43: 615–630.

McMillen, Daniel P. 1992. "Probit with spatial autocorrelation." *Journal of Regional Science*, 32(3): 335–348.

Medoff, Marshall H. 2003. "Collaboration and the quality of economics research." *Labour Economics*, 10(5): 597–608.

Miguel, Edward, and Michael Kremer. 2004. "Worms: identifying impacts on education and health in the presence of treatment externalities." *Econometrica*, 72(1): 159–217.

Muzi, Silvia, Filip Jolevski, Kohei Ueda, and Domenico Viganola. 2022. "Productivity and firm Exit during the COVID-19 Crisis: Cross-country evidence." *Small Business Economics*, 1–42.

Oettl, Alexander. 2012. "Reconceptualizing stars: Scientist helpfulness and peer performance." *Management Science*, 58(6): 1122–1140.

Osborne, Martin J, and Ariel Rubinstein. 1994. *A course in game theory.* MIT press.

Owsiak, Andrew P, and John A Vasquez. 2021. "Peaceful dyads: A territorial perspective." *International Interactions*, 47(6): 1040–1068.

Patnam, Manasa, and Jayati Sarkar. 2011. "Corporate networks and peer effects in firm policies: Evidence from India." *Department of Economics, University of Cambridge (mimeo& graph).*

Petersen, Alexander Michael. 2015. "Quantifying the impact of weak, strong, and super ties in scientific careers." *Proceedings of the National Academy of Sciences*, 112(34): E4671–E4680.

Pinkse, Joris, and Margaret E Slade. 1998. "Contracting in space: An application of spatial statistics to discrete-choice models." *Journal of Econometrics*, 85(1): 125–154.

Porter, Michael E, et al. 1998. *Clusters and the new economics of competition.* Vol. 76, Harvard Business Review Boston.

Portes, Richard, and Helene Rey. 2005. "The determinants of cross-border equity flows." *Journal of international Economics*, 65(2): 269–296.

Power, Bernadette, Justin Doran, and Geraldine Ryan. 2019. "The effect of agglomeration economies on firm deaths: A comparison of firm and regional based approaches." *Urban Studies*, 56(16): 3358–3374.

Rose, Andrew K. 2004. "Do we really know that the WTO increases trade?" *American economic review*, 94(1): 98–114.

Rosenthal, Stuart S, and William C Strange. 2001. "The determinants of agglomeration." *Journal of urban economics*, 50(2): 191–229.

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly journal of economics*, 116(2): 681–704.

Sarmiento, Camilo, and William W Wilson. 2007. "Spatially correlated exit strategies in the baking industry." *Applied Economics*, 39(4): 441–448.

Scellato, Salvatore, Cecilia Mascolo, Mirco Musolesi, and Vito Latora. 2010. "Distance matters: geo-social metrics for online social networks."

Snow, David A, Peter J Leahy, and William A Schwab. 1981. "Social interaction in a heterogeneous apartment: An investigation of the effects of environment upon behavior." *Sociological Focus*, 14(4): 309–319.

Tinbergen, Jan. 1962. "Shaping the world economy; suggestions for an international economic policy."

Ugander, Johan, Brian Karrer, Lars Backstrom, and Cameron Marlow. 2011. "The anatomy of the facebook social graph." *arXiv preprint arXiv:1111.4503*.

Vazquez-Bare, Gonzalo. 2022. "Identification and estimation of spillover effects in randomized experiments." *Journal of Econometrics*.

Vortherms, Samantha, and Jiakun Jack Zhang. 2021. "Political Risk and Firm Exit: Evidence from the US-China Trade War." *21st Century China Center Research Paper*, , (2021-09).

Wang, Hao, Alvin Chin, and Hao Wang. 2011. "Interplay between social selection and social influence on physical proximity in friendship formation."

Weterings, Anet, and Orietta Marsili. 2015. "Spatial concentration of industries and new firm exits: Does this relationship differ between exits by closure and by M&A?" *Regional Studies*, 49(1): 44–58.

Xie, Xing. 2010. "Potential friend recommendation in online social network." 831–835, IEEE.

Yang, Chao, and Lung-fei Lee. 2017. "Social interactions under incomplete information with heterogeneous expectations." *Journal of Econometrics*, 198(1): 65–83.

Appendix A

Supplementary material for Chapter 1

A.1 Overview of Foreign-invested Enterprises' Yearly Exits

Table A.1: Foreign-invested Enterprises' Yearly Exits, 2014-2021

	National		Beijing		Shanghai		Guangzhou		Shenzhen	
	Enterprises	Exits (%)	Enterprises	Exits (%)	Enterprises	Exits (%)	Enterprises	Exits (%)	Enterprises	Exits (%)
2014	307,808	14,757 (4.79%)	18,695	852 (4.56%)	53,816	1,900 (3.53%)	133,02	447 (3.36%)	29,527	1,718 (5.82%)
2015	322,537	14,810 (4.59%)	19,014	969 (5.10%)	57,076	2,183 (3.82%)	14,830	627 (4.23%)	32,468	1,410 (4.34%)
2016	345,975	15,567 (4.50%)	19,110	981 (5.13%)	59,253	2,166 (3.66%)	16,987	641 (3.77%)	39,905	1,756 (4.40%)
2017	374,505	19,813 (5.29%)	19,157	1,307 (6.82%)	61,093	3,249 (5.32%)	20,937	999 (4.77%)	48,616	1,728 (3.55%)
2018	355,244	32,673 (9.20%)	17,869	1,802 (9.52%)	57,895	5,405 (9.36%)	19,982	1,602 (8.02%)	47,010	4,619 (9.83%)
2019	322,602	29,697 (9.20%)	16,067	1,592 (9.91%)	52,500	4,973 (9.47%)	18,383	1,939 (10.55%)	42,391	4,168 (9.83%)
2020	292,921	22,600 (7.72%)	14,476	973 (6.72%)	47,529	3,101 (6.52%)	16,447	1,041 (6.33%)	38,224	2,004 (5.24%)
2021	270,333	10,188 (3.77%)	13,504	541 (4.01%)	44,428	1,418 (3.19%)	15,408	731 (4.74%)	36,222	1,435 (3.96%)

Appendix B

Supplementary material for Chapter 2

B.1 Marginal Effect of the Friendship Network

Table B.1: Marginal Effects at Mean of Logit Model in Friendship Network

	World	Thailand	Turkey	Taiwan	Indonesia	Brazil	Russia
5-10 km	-0.03^{***}	-0.03^{***}	-0.02^{***}	-0.06^{***}	-0.05^{***}	-0.01^{***}	-0.004^{***}
10-50 km	-0.04^{***}	-0.05^{***}	-0.02^{***}	-0.10^{***}	-0.08^{***}	-0.02^{***}	-0.01^{***}
50-100 km	-0.05^{***}	-0.05^{***}	-0.03^{***}	-0.13^{***}	-0.09^{***}	-0.02^{***}	-0.01^{***}
100-500 km	-0.06^{***}	-0.05^{***}	-0.03^{***}	-0.17^{***}	-0.10^{***}	-0.02^{***}	-0.01^{***}
500-1000 km	-0.06^{***}	-0.06^{***}	-0.03^{***}	-0.14^{***}	-0.10^{***}	-0.03^{***}	-0.01^{***}
1000-2000 km	-0.06^{***}	-0.06^{***}	-0.03^{***}	-0.24^{***}	-0.11^{***}	-0.02^{***}	-0.01^{***}
2000-3000 km	-0.07^{***}	-0.07^{***}	-0.03^{***}	-0.24^{***}	-0.11^{***}	-0.02^{***}	-0.01^{***}
>3000 km	-0.07^{***}	-0.07^{***}	-0.04^{***}	-0.24^{***}	-0.11^{***}	-0.03^{***}	-0.02^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

B.2 Coefficients of the Directed Network

Table B.2: Logit Coefficient of Geographical Distances in Directed Network

	World	Thailand	Turkey	Taiwan	Indonesia	Brazil	Russia
5-10 km	-0.53	-0.67***	-0.50***	-0.35***	-0.65***	-0.64***	-0.28***
		(0.01)	(0.02)	(0.02)	(0.03)	(0.04)	(0.06)
10-50 km	-1.02	-1.21***	-0.93***	-0.67***	-1.24***	-1.42***	-0.61***
		(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.05)
50-100 km	-1.37	-1.56^{***}	-1.28^{***}	-1.01^{***}	-1.63^{***}	-1.96^{***}	-0.92^{***}
		(0.01)	(0.02)	(0.02)	(0.03)	(0.04)	(0.07)
100-500 km	-1.71	-1.88^{***}	-1.71^{***}	-1.44^{***}	-2.22^{***}	-2.26^{***}	-0.96^{***}
		(0.01)	(0.01)	(0.02)	(0.03)	(0.04)	(0.06)
500-1000 km	-1.93	-1.97^{***}	-1.88^{***}	-1.17^{***}	-2.62^{***}	-2.41^{***}	-1.57^{***}
		(0.01)	(0.01)	(0.09)	(0.03)	(0.04)	(0.06)
1000-2000 km	-2.40	-2.50^{***}	-1.90^{***}	-5.18^{***}	-2.86^{***}	-2.57^{***}	-1.67^{***}
		(0.02)	(0.02)	(0.15)	(0.03)	(0.05)	(0.05)
2000-3000 km	-2.89	-3.10^{***}	-1.98^{***}	-5.25^{***}	-3.49^{***}	-1.99^{***}	-1.77^{***}
		(0.03)	(0.02)	(0.07)	(0.04)	(0.05)	(0.06)
>3000 km	-3.55	-3.21^{***}	-2.83^{***}	-6.69^{***}	-4.04^{***}	-3.31^{***}	-1.86^{***}
		(0.03)	(0.03)	(0.09)	(0.04)	(0.08)	(0.06)

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.3: Logit Coefficient of Homophily Variables in Directed Network

	World	Thailand	Turkey	Taiwan	Indonesia	Brazil	Russia
Country Popularity (i)	-0.21						
Country Popularity (j)	-0.31	-0.41***	-0.33***	0.78***	-0.39***	-0.05	0.00
		(0.03)	(0.02)	(0.05)	(0.04)	(0.04)	(0.03)
Same Platform	0.10	0.06***	0.07***	0.05***	0.09**	0.11***	0.12***
		(0.01)	(0.01)	(0.01)	(0.04)	(0.02)	(0.01)
Same Continent	-0.35	-0.08	-0.04	-7.49	0.26***	0.27***	-0.31***
		(0.10)	(0.05)	(48.31)	(0.05)	(0.08)	(0.05)
Same Country	1.70	0.66***	0.55***	1.44***	0.60***	-0.03	0.35***
		(0.08)	(0.06)	(0.06)	(0.04)	(0.12)	(0.04)
Same Region	0.24	0.93***	0.45***	0.08***	0.20***	0.52***	0.50***
		(0.03)	(0.05)	(0.01)	(0.02)	(0.02)	(0.08)
Same City	0.07	-0.89***	-0.05	0.40***	0.09***	0.12***	-0.20**
		(0.03)	(0.05)	(0.06)	(0.02)	(0.02)	(0.08)
Follower Younger	-0.16	-0.06***	-0.24**	-0.03**	0.02	-0.08***	-0.09***
		(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Age Within 5 Years	0.14	0.13***	0.17***	0.26***	0.08***	0.09***	0.18***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Same Ethnicity	0.16	0.04***	0.00	0.07***	0.02	0.05***	0.02
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Follower Taller	0.03	0.02*	0.05***	-0.06***	-0.17***	-0.05***	0.02
		(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Height Within 5cm	0.07	0.05***	0.06***	0.04***	0.16***	0.06***	0.06***
		(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)
Follower Heavier	0.04	0.03***	0.02*	0.04**	0.01	-0.02	-0.05**
		(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)
Weight Within 5kg	0.01	0.02	0.01	-0.04	-0.08	0.01	0.01
		(0.02)	(0.03)	(0.03)	(0.07)	(0.05)	(0.04)
Same Sex Role	-0.27	-0.17***	-0.68***	-0.21***	-0.07***	-0.11***	-0.17***
		(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)
Same Language	0.42	0.23***	0.40***	0.20***	0.34***	0.10**	0.06
		(0.01)	(0.02)	(0.04)	(0.02)	(0.05)	(0.06)

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.4: Logit Coefficient of Control Variables in Directed Network - Part 1

	World	Thailand	Turkey	Taiwan	Indonesia	Brazil	Russia
Posts Made in V4 (i)	−0.03	−0.02***	−0.08	−0.55	0.46***	−1.20***	−3.74***
		(0.00)	(0.07)	(0.70)	(0.07)	(0.39)	(0.41)
Posts Made in V5 (i)	0.00	−0.01**	−0.10***	0.03**	−0.01***	0.17***	0.04***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Posts Made in V6 (i)	−0.03	−0.01*	−0.00	−0.02	−0.04***	−0.00	−0.09***
		(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.02)
Posts Liked/Commented in V4 (i)	−0.00	−0.02***	0.23***	0.00	−0.05***	−0.08***	0.00
		(0.01)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)
Posts Liked/Commented in V5 (i)	0.07	0.00	−0.04***	−0.10***	0.04***	0.15***	0.07***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Posts Liked/Commented in V6 (i)	0.03	−0.02***	0.14***	0.22***	0.03***	0.02***	−0.02***
		(0.00)	(0.00)	(0.02)	(0.01)	(0.00)	(0.00)
Stories Read from Feed in V4 (i)	0.02	−0.03***	0.03***	−0.03**	0.04**	0.19***	0.17***
		(0.01)	(0.00)	(0.01)	(0.02)	(0.01)	(0.01)
Stories Read from Feed in V5 (i)	−0.01	0.02***	−0.05***	−0.03**	0.02***	−0.05***	0.04***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Stories Read from Feed in V6 (i)	−0.00	−0.00	−0.12***	−0.12***	0.01**	−0.04***	−0.04***
		(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)
Chat Messages Sent in V4 (i)	0.01	0.01***	−0.02***	0.03***	0.06***	0.20***	−0.03***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Chat Messages Sent in V5 (i)	−0.00	0.01***	0.02***	−0.02***	0.06***	0.21***	0.08***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Chat Messages Sent in V6 (i)	0.06	−0.01***	0.07***	0.01*	0.07***	0.03***	−0.05***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Total Followees in V4 (i)	0.01	0.02***	−0.04***	−0.09***	0.05***	−0.79***	0.00
		(0.00)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
Total Followees in V5 (i)	−0.16	0.06***	0.09***	−0.09***	−0.15***	−0.29***	−0.15***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Total Followees in V6 (i)	−0.07	−0.03***	−0.06***	−0.00	−0.04***	−0.07***	0.08***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Posts Made in V4 (j)	0.00	0.01***	−0.01	0.01	−0.02***	−0.01	−0.02*
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Posts Made in V5 (j)	0.01	0.02***	0.02***	0.02***	0.02***	0.02***	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Posts Made in V6 (j)	0.00	−0.04***	−0.02***	−0.05***	0.05***	0.03***	0.03***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)

Note: *p<0.1; **p<0.05; ***p<0.01

Table B.5: Logit Coefficient of Control Variables in Directed Network - Part 2

	World	Thailand	Turkey	Taiwan	Indonesia	Brazil	Russia
Posts Liked/Commented in V4 (j)	−0.00	0.02***	0.00	−0.02**	−0.01	0.00	−0.01
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Posts Liked/Commented in V5 (j)	0.02	0.00**	0.03***	0.03**	0.04***	0.02***	0.07***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Posts Liked/Commented in V6 (j)	0.03	0.02***	0.04***	0.02	0.03***	0.03***	−0.01***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Stories Read from Feed in V4 (j)	−0.00	0.01***	−0.01***	−0.05***	−0.01*	−0.01**	−0.02***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Stories Read from Feed in V5 (j)	−0.01	−0.02***	−0.01***	0.03***	−0.00	−0.04***	0.01*
		(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
Stories Read from Feed in V6 (j)	0.02	−0.00	0.01***	−0.06***	0.02***	0.03***	−0.00
		(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
Chat Messages Sent in V4 (j)	0.01	−0.01***	0.01**	−0.04***	0.05***	0.01**	0.04***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Chat Messages Sent in V5 (j)	−0.01	0.01***	−0.07***	0.08***	−0.06***	0.02***	0.03***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Chat Messages Sent in V6 (j)	0.07	0.04***	0.09***	0.13***	0.09***	0.02***	0.05***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total Followers (i)	0.09	0.02***	−0.03***	0.02**	0.01	−0.07***	0.02**
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Total Followees (i)	0.44	0.53***	0.47***	1.75***	0.99***	0.84***	0.96***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Total Followers (j)	0.41	0.39***	0.42***	0.57***	0.38***	0.47***	0.42***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total Followees (j)	0.05	0.05***	0.07***	−0.07***	0.08***	0.01*	0.03***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix C

Supplementary material for Chapter 3

C.1 Additional Assumption for Parameter Identification

Assumption 7. *(i) For any i, ϵ_i is independent of $\mathbf{Z}$ and $\mathbf{G}$;*

(ii) $\sup_i \sum_{j=1}^{n} \mathbf{g}_{ij}$ is uniformly bounded in n. The element $\mathbf{g}_{ij}$'s are at most of order h_n^{-1} uniformly in all i, j, where the sequence $(h_n)_n$ can be bounded or divergent, such that h_n/n converges to 0 as n grows to infinity;

(iii) supp(ω) is not contained in a proper linear subspace of $\mathbb{R}^{K+1}$. If a subgroup s contains a positive proportion of individuals who have peers, then supp$(\tilde{\omega}_{|s})$ is not contained in a proper linear subspace of $\mathbb{R}^{2K+1}$, where $\tilde{\omega}_{|s}$ is $\tilde{\omega}_i$ for an arbitrary i from the subgroup s;

(iv) There exists $k_0 \in [1, K]$ such that $\beta_{k_0}\gamma_{k_0} \geq 0$ and $\gamma_{k_0} \neq 0$. There exists a subgroup s_0 in which the proportion of agents who have peers is strictly positive, also as n grows to infinity. The cardinality of the set $T_{s_0,n} = \{i \in s_0 : \exists j, l \in s_0, \text{where} i \neq l \text{such that} \mathbf{g}_{ij} > 0, \mathbf{g}_{jl} > 0, \mathbf{g}_{il} = 0\}$ in proportion to $|s_0|$ is strictly positive, also as n grows to infinity.

C.2 Contextual Effect on Scholars' Number of Publications

Table C.1: Contextual Effects on Scholars' Number of Publications in Non-Covid and Covid Times

	Pre-Covid (2018-2019)	Covid (2020-2021)	Covid + Covid Index (2020-2021)
Proportion of Coauthors with 2-4 Publications Per Year	0.05	0.13	0.11
	(0.23)	(0.17)	(0.17)
Proportion of Coauthors with 5-9 Publications Per Year	−0.31	0.31	0.28
	(0.33)	(0.32)	(0.32)
Proportion of Coauthors with 10+ Publications Per Year	−0.59	−0.27	−0.34
	(0.57)	(0.54)	(0.54)
Proportion of Coauthors with 100-499 Citations	0.11	0.44*	0.46*
	(0.26)	(0.27)	(0.27)
Proportion of Coauthors with 500-1,999 Citations	−0.02	0.37*	0.36
	(0.21)	(0.20)	(0.20)
Proportion of Coauthors with 2,000-4,999 Citations	−0.07	−0.09	−0.07
	(0.35)	(0.30)	(0.30)
Proportion of Coauthors with 5,000-9,999 Citations	−0.20	0.15	0.16
	(0.41)	(0.38)	(0.37)
Proportion of Coauthors with 10,000-19,999 Citations	−0.49	0.26	0.28
	(0.37)	(0.41)	(0.41)
Proportion of Coauthors with 20,000+ Citations	−0.52	0.38	0.38
	(0.58)	(0.53)	(0.53)
Proportion of Coauthors with 10-20 Years Experience	0.22	0.15	0.16
	(0.30)	(0.22)	(0.22)
Proportion of Coauthors with 20-30 Years Experience	−0.39	0.05	0.05
	(0.33)	(0.24)	(0.24)
Proportion of Coauthors with 30-40 Years Experience	−0.39	−0.28	−0.26
	(0.35)	(0.28)	(0.28)
Proportion of Coauthors with 40-50 Years Experience	−0.02	−0.69**	−0.69**
	(0.37)	(0.32)	(0.32)
Proportion of Coauthors with 50-60 Years Experience	−0.14	−0.49	−0.48
	(0.42)	(0.39)	(0.39)
Proportion of Coauthors with 60+ Years Experience	−1.09**	0.48	0.50
	(0.43)	(0.40)	(0.40)
Covid Index (Coauthors)			−0.35
			(0.45)
African American (Coauthors)	0.12	−0.68	−0.68
	(0.52)	(0.78)	(0.77)
Female (Coauthors)	0.08	0.01	0.02
	(0.19)	(0.19)	(0.19)
Field: Theory (Coauthors)	−0.06	−0.10	−0.12
	(0.21)	(0.17)	(0.17)
Field: Macro (Coauthors)	−0.01	0.35**	0.33**
	(0.15)	(0.16)	(0.16)
Field: Labor (Coauthors)	−0.45**	0.15	0.16
	(0.21)	(0.16)	(0.16)
Field: Metrics (Coauthors)	−0.06	−0.07	−0.07
	(0.17)	(0.17)	(0.17)
Field: Industrial Organization (Coauthors)	−0.32	0.16	0.16
	(0.23)	(0.27)	(0.26)
Field: Development (Coauthors)	−0.20	−0.04	−0.03
	(0.19)	(0.20)	(0.20)
Field: Health (Coauthors)	0.33	0.05	0.07
	(0.27)	(0.28)	(0.28)
Field: Finance (Coauthors)	−0.03	−0.10	−0.10
	(0.21)	(0.21)	(0.21)

Note: *p<0.1; **p<0.05; ***p<0.01