UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on the Identification and Estimation of Network Models

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Yiran Xie

2022

ABSTRACT OF THE DISSERTATION

Essays on the Identification and Estimation
of Network Models

by

Yiran Xie
Doctor of Philosophy in Economics
University of California, Los Angeles, 2022
Professor Rosa Liliana Matzkin, Chair

This dissertation consists of three main chapters that study social interactions in networks. In Chapter 1, I study a market with many-to-many contracts when the number of market participants increases. Many-to-many contracts allow a seller to trade with multiple buyers and a buyer to trade with multiple sellers. I focus on investigating the identification of payoff parameters through data observed from equilibrium matches in a large many-to-many matching market. In many-to-many matching markets, several issues have to be addressed: bias would arise since the outcomes are only observed when links are formed between two agents, and the maximum number of relationships an agent can enter into would possibly affect the set of stable outcomes. I show that under certain conditions, the number of firms (workers) that are willing to be matched to a specific worker (firm) grows at a rate regardless of the capacity of both sides. Furthermore, I show a correspondence between the stable matching outcomes in a many-to-many matching framework and that in a one-to-one matching framework.

In Chapter 2, I conduct a structural econometric analysis of the diffusion process with players who observe their neighbors and make decisions based on their neighbors' decisions. I study the identification and estimation of diffusion processes in social and economic net-

works. Compared to the classic econometric diffusion literature that assumes a continuous population with a stochastic network structure, I provide a new econometric framework to analyze diffusion processes in fixed networks where Bayesian players observe their close neighbors. I demonstrate the existence of the equilibrium of the model and characterize the unique symmetric equilibrium. Based on these theoretical findings, I propose a consistent and tractable two-step estimator for payoff parameters using feasible data from a single large network. I evaluate the finite sample performance using Monte Carlo simulations.

Chapter 3 applies the network diffusion model to data on the participation of a microfinance program in Indian villages to describe the impact of neighbors on individual decisions. Our model allows us to study the various network effect across different types of agents who care about their neighbors' opinions. It depends on unknown equilibrium beliefs, which specify agents' expectations about their neighbors' decisions. Using participation data from 43 villages, each including about 200 villagers, I estimate these equilibrium beliefs and the network effects.

The dissertation of Yiran Xie is approved.

Hyungsik Moon

Denis Nikolaye Chetverikov

Shuyang Sheng

Andres Santos

Rosa Liliana Matzkin, Committee Chair

University of California, Los Angeles

2022

To my parents . . .

for their love and support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisors, Rosa Matzkin, Andres Santos, Shuyang Sheng, for their invaluable patience, guidence and continuous support. This endeavor would not have been possible without my committee, Denis Chetverikov and Hyungsik Moon, who generously provided knowledge and expertise. Thanks should also go to Jinyong Hahn, Zhipeng Liao and other econometrics seminar participants at the University of California, Los Angeles.

I am grateful to my family, especially my parents and spouse. Their belief in me has kept my spirits and motivation high during this process. I would also like to recognize Zhu Yilong for all the moral and emotional support.

VITA

Education

2018	C.Phil., Economics, Department of Economics
2018	M.A. in Economics, University of California, Los Angeles (UCLA)
2016	B.A. in Economics, Guanghua School of Management, Peking University (PKU)
2016	B.S. in Applied Mathematics (double major), Peking University (PKU)

Awards

2021	Best Paper Award, Econometrics Proseminar
2020	Summer Research Fellowship, UCLA Economics
2018	Cathay Bank Scholarship
2016-2021	University Fellowship, UCLA
2016	Outstanding Graduate Student, Peking University
2013-2015	Fengqi Scholarship, POSCO Scholarship, Yihai Kerry Scholarship, Peking University
2011	Best Female Participant and Gold Medal, 28$^{\text{th}}$ Chinese Physics Olympiad (CPhO)

CHAPTER 1

Identification in Many-to-Many Matching Games

1.1 Introduction

Consider a collection of firms and workers that try to form employer-employee relationship. Firms have preferences over the possible sets of workers, and workers have preferences over the possible sets of firms. A many-to-many matching is an assignment of sets of workers to firms, and of sets of firms to workers so that each agent can be assigned to multiple agents from the other group.

There are many important real-world many-to-many markets. For example, a physician can hold multiple appointments at different institutions, that is, while having a medical position in a hospital, he or she could also teach in a university and work for private consultation. Roth (2003)[49] discusses the market for medical interns in the U.K. The medical interns are required to experience both medical and surgical positions, so the market needs to be modeled as a two-sided many-to-many matching problem where the capacity of a medical intern is 2. Many-to-many matching also exists in the model of contracting between down-stream firms and up-stream providers, or between customers and products. A customer can purchase various products, and a product can be purchased by many customers.

In a market where many-to-many matching is allowed, the equilibrium result can be very different compared to the market with only one-to-one matching. Echenique and Oviedo (2006)[19] shows that even a few many-to-many contracts can make a crucial difference. In the real world, most labor markets have at least a few many-to-many contracts. Thus one needs a many-to-many model to study the labor markets.

In this chapter, I study the identification of payoff parameters with data observed from equilibrium matches in a large many-to-many matching market. The payoff parameters are key to identify the production function and the surplus from an equilibrium match. This chapter focuses on the case of a bipartite graph. Consider the example of the firm-worker relationship. The payoff a worker i gets from being employed by firm j is given by

$$U_{ij}^* = \tilde{U}(x_i, z_j) + \sigma\eta_{ij} \overset{(e.g.)}{=} x_i\alpha + z_j\beta + \sigma\eta_{ij},$$

where X_i, Z_j consists of the observable characteristics of worker i and firm j. η_{ij} is the idiosyncratic shock of the payoff to each pair (i, j). U^*_{ij} is only observable when there is a link formed between i and j - that is, when worker i is employed by firm j - and the link formation depends on the payoffs of all players in i and j's opportunity sets.

It's worth noticing that this model requires no high-level assumptions. Due to the complexity of the matching mechanism, high-level assumptions, for example, rank-order property, which is prevalent in the matching literature. Moreover, my model is related to literature that studies models with endogenous link formation, such as Johnsson and Moon (2021)[36] and Auerbach (2019)[2]. Unlike in their settings - where the probability that two agents link is determined by some unknown function of their own social characteristics, thus not affected by other agents in the network formation game - my model has agents compare the payoffs from linking with every player who is willing to connect and form links only when the payoff of the matching is among the highest in the choice set. The link formation depends not only on the characteristics of players on the two ends of the link, but also on the opportunity set that characterizes the players who are willing to connect.

To investigate the identification of the payoff parameters, the paper demonstrate the correspondence between the many-to-many matching and one-to-one matching following Menzel (2015)[43]. We show that the conditional expectation of payoff of a matching, given the link is observed, depends only on the inclusive value. Thus even if we observe the payoffs from workers, we still cannot separately identify the deterministic part of the payoff that depends only on the observable characteristics.

The contribution of the paper is threefold. First, it builds its discussions on a many-to-many matching framework that requires no high-level assumptions, which has not been investigated before. The complexity of matching mechanisms arise with the absence of such assumptions, but it is achievable and beneficial- in that it clarifies how the rank-order property is driven by the basic assumptions.

Furthermore, we demonstrate that the many-to-many matching outcomes correspond to one-to-one matching outcomes. As illustrated in the paper, the Gale-Shapley algorithms[1] yield setwise-stable matchings under the max-min preferences, and the number of links a player could form - though it affects the density of matchings over types - would not affect the growth rate of the size of the opportunity set. Therefore, we can connect the results of the many-to-many matching framework to that of the one-to-one matching framework, as the inclusive value is very similar.

Lastly, we show that the inclusive values can be identified using information about the payoff from one side of the market, but it cannot help separately identify the payoff of players from the total welfare. In the many-to-many matching framework, we prove that the payoff of a worker, given it is observed, depends on the attributes of all firms in her opportunity set and is characterized by the inclusive values. Importantly, the payoff does not depend on the specific characteristics of the firm she connects to.

1.1.1 Related Work

A strand of literature studies matching models with transferable utility. A first group of methods restrict the distribution of the unobserved heterogeneity ε_{ij}. Choo and Siow (2006b)[13] assumes the surplus function is additively separable in the unobserved components of both partners. Galichon and Salanié (2015)[24] show that given exact knowledge of the parametric specification of the stochastic terms, the mean joint surplus is nonparametrically just identified. Graham (2011, 2013) shows sign-based identification assuming that unobservables are independently and identically distributed. All these papers cannot separately identify utility functions from the two sides when the transfers are not observed.

In contrast, Fox (2010)[22] has proposed an approach that does not explicitly specify the distribution of the unobserved heterogeneity. Instead, it directly postulates a *rank-order property* that imposes restrictions on the relationship between matching patterns and the surplus function. Two sets are matched more frequently if they have higher joint surplus. However, this paper is based on the high level assumption (rank-order property) and such

[1] Also known as the Deferred Acceptance algorithm.

monotonic behavior is rare in matching models. But it is worth stressing here that because Fox's approach only relies on pairwise stability, it can be applied more widely, to many-to-one or even many-to-many matching.

For matching models with nontransferable utility, Hsieh (2012) assumes each agent equally values his or her potential partners who belong to the same category. Menzel (2015)[43] yields a remarkably simple asymptotic formula of one-to-one matching with nontransferable utility, but the model also cannot separately identify the utility functions. Diamond and Agarwal (2017)[17] assume that preferences are homogeneous and the attractiveness can be summarized using a single index. That is, each side of the market is only vertically differentiated.

A group of related literature studies models with endogenous link formation. Johnsson and Moon (2015)[36] and Auerbach (2019)[2] both consider a model of link formation in which the probability that two agents link is some unknown function of their social characteristics $D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}$. In my model setup, the outcomes are also only observed when links are formed between the two agents. But I consider the case that an agent compares payoffs of every agent in his or her choice set, since links are formed only when payoffs of the matching is among the K highest in the choice set. The link formation depends not only the social characteristics of the two parties, but also social characteristics of those who are available to them.

The framework of many-to-many matching has been studied in the theory literature. Early papers such as Roth (1984)[48] proposes *setwise-stability* as an equilibrium concept in many-to-many matching, and Sotomayor (1999)[55] emphasizes the difference between setwise-stability, pairwise-stability, and the core. For many-to-many matching, although the Gale-Shapley algorithms continue to yield pairwise-stable matchings, this outcome may no longer be setwise-stable in a many-to-many matching problem. There can be a group deviation from a pairwise-stable matching that improves the payoff of every member of the deviation. Two more recent papers show group deviation from a matching is not executable under certain conditions. Echenique and Oviedo (2006)[19] gives conditions under which the setwise-stable set is nonempty and can be approached through an algorithm. Konishi and

Ünver (2005)[37] is an independent work shows that a concept they call credible group stability is equivalent to pairwise stability in a wide class of matching problems when the preferences are responsive. Jiao and Tian (2015)[35] demonstrate the equivalence of setwise-stability and pairwise-stability obtained under max-min preference.

1.2 Model

Consider an undirected unweighted graph G in which there are two sets of agents, F and W. We focus on the case of a bipartite graph, that is, the two sets F and W do not overlap. Suppose links are formed only between the subsets F and W. So for a link between a pair of agents (i, j), we have that $i \in F$ and $j \in W$. The graph describes the interaction between the two types of units, such as workers and firms or students and teachers. The corresponding outcome of interest are wages or test scores.

There are n_F firms and n_W workers in the market. The agents have a quota K giving the maximum number of partnerships it may enter into, K_F for firms and K_W for workers. Assume $n_F = O(n^\alpha)$, $K_F = O(n^{1-\alpha})$, $n_W = O(n)$, and $K_W = O(1)$.

Agents maximize their total utility from all their links. Firms' outcomes U and workers' outcomes V over matchings depend on their characteristics (X_i, Z_j) and idiosyncratic shocks η_{ij} and ξ_{ji}.

$$U^*_{ij} = \tilde{U}_{ij} + \sigma\eta_{ij} \overset{(e.g.)}{=} X_i\alpha_F + Z_j\beta_F + \sigma\eta_{ij},$$

$$V^*_{ji} = \tilde{V}_{ji} + \sigma\xi_{ji} \overset{(e.g.)}{=} X_i\alpha_W + Z_j\beta_W + \sigma\xi_{ji},$$

for $i = 1, ..., n_F$ and $j = 1, ..., n_W$.

Although the capacity is fixed, firms (workers) do not necessarily have to connect to K_F workers (F_W firms) since we assume there are J outside options for both groups. The outcomes of the outside options are

$$U^*_{iO_1} = 0 + \sigma\eta_{iO_1}, ...U^*_{iO_J} = 0 + \sigma\eta_{iO_J},$$

$$V_{jO_1}^* = 0 + \sigma\xi_{jO_1}, ..., V_{jO_J}^* = 0 + \sigma\xi_{jO_J},$$

where J increases at a rate to be specified later. The idiosyncratic component η_{ij}, ξ_{ji} and η_{iO_l}, ξ_{jO_l} are i.i.d. from a standard normal distribution, independent of $\{X_i, Z_j\}$. The outside options in a firm-worker example can possibly be time for leisure, rest, or spending time with family. An agent considers both inside and outside options. They compare the payoffs they get from connecting with potential partners in the market, and also the payoffs of connecting to outside options. The outcome is not observable if the link connects the agent to an outside option.

We observe data (A_{ij}, X_i, Z_j) for each i, j and U_{ij} only when a link is formed - in the firm-worker example, the wage is only observed when agent pairs enter into a firm-worker relationship - where

$$A_{ij} = A_{ji} = \mathbb{1}\{i \in F_j, j \in W_i\} = \mathbb{1}\{U_{ij}^* \geq U_{im_{(K_F)}}^*, V_{ji}^* \geq V_{jn_{(K_W)}}^*\}$$

$$U_{ij} = U_{ij}^* \cdot A_{ij}$$

and $A_{ij} = 1$ if a link forms between agent i and j.

Following Menzel (2015)[43], the rationale for allowing the agent to sample an increasing number J of independent draws for the outside option is that, since the shocks η_{ij} and ξ_{ji} have unbounded support, as the set of potential matching partners grows with number of agents n, any alternative with a fixed utility level will eventually be dominated by the K largest draws. Hence, we assume that as the market grows, the typical agent can choose from an increasing number of potential matchings, thus the outside options are sufficiently attractive to ensure that the share of agents who choose the outside option does not degenerate. In later sections, I show that J should grow in the same rate $(\sqrt{n})$ as the size of the set of potential matching partners.

Assumption 1. *(Idiosyncratic Part of Payoffs) η_{ij} and ξ_{ji} are i.i.d. draws from the distribution $G(s)$, and are independent of x_i, z_j, where* *(i) the c.d.f. $G(s)$ is absolutely continuous with density $g(s)$,*

(ii) the upper tail of the distribution $G(s)$ is of type I with auxiliary function $a(s) := \frac{1-G(s)}{g(s)}$.

Assumption 1 provides sufficient conditions for the distribution of η_{iO_l} to belong to the domain of attraction of the extreme-value type-I distribution. Assumption 1 requires the taste shifter is extreme-value type-I, which causes the conditional choice probability converges to a logit result.

Assumption 2. *(Market Size)*

(i) The size of a given market is governed by $n = 1,2,\ldots$ with the number of firms and workers $n_F = n^{\alpha} \exp \gamma_F$ and $n_W = n$;

(ii) the size of the capacity of firms and workers: $K_F = n^{1-\alpha} \exp \gamma'_F$ and $K_W = \exp \gamma'_W$ is a fixed number.

(iii) the scale parameter for the unobservables $\sigma \equiv \sigma_n = \frac{1}{a(b_n)}$, where $b_n = G^{-1}(1 - \frac{1}{\sqrt{n}})$, and $a(s)$ is the auxiliary function in Assumption 1.

Assumption 2 gives the approximating sequence of markets. Specifically, we want the approximation to keep several qualitative features that we observe in the finite-agent market:

First, the share of links that have one end connect to the outside option should not degenerate to 1 or zero so that agents would be always able to choose among the inside and outside option. Thus it is necessary to increase the payoff from outside option as the number of available alternatives grows. This is governed by an assumption on the size of the set of outside options we will see later.

Second, we want the systematic parts of payoffs to remain predictive for match probabilities in the limit. Thus we have to choose the scale parameter $\sigma \equiv \sigma_n$ at an appropriate rate to balance the relative scales of the systematic and idiosyncratic parts. This is achieved by the third part of assumption 1. It is also an assumption used in Menzel (2015)[43].

1.3 Setwise-stable Matching from Max-min Utility

In the setting, agents maximize the sum of the utility from all their links. This could instead be characterized by the max-min utility that denotes the preference that agents choose to form links with the candidates who bring the highest K payoffs. Denote the outside options by $O = \{O_1, ..., O_J\}$. Agents compare payoffs of all elements in $S \cup O$. Define max-min utility of firm i as

$$U_i(S) = \max_{|K|=K_F} \min_{j \in K} U_{ij} = U_{i(K_F)}$$

where $K \subseteq S \cup O$, and $U_{i(K_F)}$ is the K_F-th largest utility.

Max-min maximizes the minimum utility in the choice set. It satisfies strong substitutability, which is a condition that leads to the setwise-stability. Jiao and Tian (2015) show the equivalence of setwise-stability and pairwise-stability obtained under max-min preference. Thus the setwise-stable matching can be achieved by algorithms that generates a pairwise-stable matching.

1.4 Asymptotic Properties

1.4.1 The Correspondence between the Many-to-Many Matching and One-to-One Matching

In this section, I demonstrate the correspondence between the many-to-many matching in my model and the one-to-one matching in Menzel's setting.

To develop the asymptotic argument, the first step is to derive the convergence of conditional choice probabilities (CCP) to logit CCPs, under the assumption that unobservables η_{ij} are independent from the equilibrium opportunity sets W_i and M_j, where the CCP denotes the probability of firm i prefers worker j from the set of worker 1 to J, conditional on firm i's utility when it hires worker 1 to J.

Lemma 1. *Suppose that Assumption 1 and 2 hold, and the random utilities $U_{i1}, ..., U_{iJ}$ are*

J i.i.d. draws from the model with J outside options. Then as $J \to \infty$, the marginal CCP,

$$\left| JP(i \in F_j | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{K_F \exp\{\tilde{U}_{ij}\}}{1 + \frac{1}{J}\sum_{k=1}^{J} \exp\{\tilde{U}_{ik}\}} \right| \to 0$$

Similarly,

$$\left| JP(j \in W_i | \tilde{V}_{j1}, ..., \tilde{V}_{jJ}) - \frac{K_W \exp\{\tilde{V}_{ji}\}}{1 + \frac{1}{J}\sum_{k=1}^{J} \exp\{\tilde{V}_{jk}\}} \right| \to 0$$

where F_j is the set of firms available to worker j, and W_i is the set of workers available to firm i. $\tilde{U}$'s and $\tilde{V}$'s are the systematic part of utility, $\tilde{U}_{ij} = U(x_i, z_j)$.

Proof. See Appendix. $\square$

Recall that in Menzel (2015)[43],

$$\left| JP(U_{ij} \geq U_{ik}, k = 1, ..., J | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{\exp\{\tilde{U}_{ij}\}}{1 + \frac{1}{J}\sum_{k=1}^{J} \exp\{\tilde{U}_{ik}\}} \right| \to 0$$

Thus Lemma 1 illustrates the correspondence when the capacity is fixed. As we will show in the later sections that J grows with rate $\sqrt{n}$, the empirical matching frequencies of this many-to-many matching follows

$$f(x, z) = \frac{K_F K_W \exp\{U(x, z) + V(z, x) + \gamma_F + \gamma_W\} f(x) w(z)}{(1 + \Gamma_F(x))(1 + \Gamma_W(z))}$$

1.4.2 Size of opportunity sets F_j and W_i: $O(\sqrt{n})$

In this section, I show the size of the opportunity set is $O(\sqrt{n})$, regardless of the fact that K_F grows with rate $O(n^{1-\alpha})$ and n_F grows with rate $O(n^{\alpha})$.

Proposition 1. *(Size of Opportunity Set) Suppose Assumption 1 and 2 hold, then the size of the opportunity set grows with rate $\sqrt{n}$.*

Proof. See Appendix. $\square$

Therefore, we than have the expected inclusive value function in the following fixed-point characterization:

Theorem 1. *(Inclusive Values) Suppose Assumption 1 and 2 hold.*

$$\hat{\Gamma}_f(x) = \frac{K_W}{n}\sum_{k=1}^{n_W}\frac{\exp\{U(x,z_k)+V(z_k,x)\}}{1+\hat{\Gamma}_w(z_k)}$$
$$\hat{\Gamma}_w(z) = \frac{K_F}{n}\sum_{k=1}^{n_F}\frac{\exp\{V(z,x_k)+U(x_k,z)\}}{1+\hat{\Gamma}_f(x_k)}$$

Proof. See Appendix. □

Notice that as K_F grows with rate $O(n^{1-\alpha})$ and n_F grows with rate $O(n^{\alpha})$, the inclusive value functions could converges in the limit.

Menzel (2015)[43] shows that the inclusive value function is point-identified.

1.4.3 Conditional Expected Utility of Observed Matchings

In this section, we demonstrate the result that the payoff observed from one side of the market (e.g., wage in the firm-worker matching market) helps identify the inclusive value function, but could not provide any information about the deterministic part of the payoff $\tilde{U}_{ij}$.

Suppose player i makes decision to connect over a set M and J outside options. Suppose $\#M = J$. Let $T = \sigma^{-1}\left(U_{im_{(K_F+1)}} - \tilde{U}_{ij}\right) = \eta_{im_{(K_F+1)}} + \sigma^{-1}\left(\tilde{U}_{im_{(K_F+1)}} - \tilde{U}_{ij}\right)$, where $m_{(K_F+1)}$ denotes the (K_F+1)-th largest element in the set M. We can show that the following result.

Lemma 2. *Under assumption 1 , T follows a Gumbel distribution with $\mu = \Phi^{-1}(1-1/J) + \sigma^{-1}\log C$ and $\beta = \sigma^{-1}$, conditional on $x_i, (z_j)_{j\in M}$.*

Theorem 2. *Suppose that Assumption 1 holds, and the random utilities $U_{i1}, ..., U_{iJ}$ are J i.i.d. draws from the model with J outside options. Then as $J \to \infty$,*

$$E[U_{ij}|X, Z, A_{ij} = 1] \to \log C + 1 + \kappa$$

where

$$C = 1 + \frac{1}{J}\sum_{m=1}^{J}\exp\{\tilde{U}_{im}\},$$

and $\gamma \approx 0.5772$ *is the Euler's constant.*

Note that $\frac{1}{J}\sum_{m=1}^{J}\exp\{\tilde{U}_{im}\}$ converges to the inclusive value when $J \to \infty$ and when M is the opportunity set.

By Proposition 1, when we take $J = \sqrt{n}$, the expected value of the maximum payoffs is depends on the inclusive value function and thus the inclusive value function can be point-identified. However, the expected maximum payoff of works with characteristic x_i does not depend on the specific characteristic of the firm z_j. This illustrates that we cannot separately identify the utility from one size of the market. Any equilibrium matching, in this setting, depends on the total surplus.

1.5 Conclusion

This paper demonstrates the correspondence in outcomes for stable matchings in a many-t-many framework and that in a one-to-one framework. We show that even if the capacity K grows with the size of the market, the size of the opportunity set remains at the same rate of growth ($\sqrt{n}$) as in the one-to-one framework. The inclusive value function can be easily identified if payoff from one side of the market is observe, but we still cannot identify the payoff separately from the total welfare.

For future research, we can look into the case that the capacity is endogenous, as the outcome from endogeneity contains information about the preference.

1.6 A. Proofs of Main Results

Proof of Lemme 1. When $K = 2$ for both firms and workers,

$$\Bigg| J^2 P(U_{ij_1} \geq U_{ik}, U_{ij_2} \geq U_{ik}, k \in \{1, ...J\} \backslash \{j_1, j_2\} | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{2\exp\{\tilde{U}_{ij_1} + \tilde{U}_{ij_2}\}}{\left(1 + \frac{1}{J}\sum_{k=1}^{J}\exp\{\tilde{U}_{ik}\}\right)^2} \Bigg| \to 0$$

$$\Bigg| JP(U_{ij} \geq U_{ik}, U_{iO_{(1)}} \geq U_{ik}, k = 1, ..., J | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{2\exp\{\tilde{U}_{ij}\}}{\left(1 + \frac{1}{J}\sum_{k=1}^{J}\exp\{\tilde{U}_{ik}\}\right)^2} \Bigg| \to 0$$

$$\Bigg| P(U_{iO_{(1)}} \geq U_{ik}, U_{iO_{(2)}} \geq U_{ik}, k = 1, ..., J | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{1}{\left(1 + \frac{1}{J}\sum_{k=1}^{J}\exp\{\tilde{U}_{ik}\}\right)^2} \Bigg| \to 0$$

Thus the "marginal" CCP follows,

$$\left| JP(U_{ij} \geq U_{ik,(2)}, k = 1, ..., J | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{2\exp\{\tilde{U}_{ij}\}}{1 + \frac{1}{J}\sum_{k=1}^{J}\exp\{\tilde{U}_{ik}\}} \right| \to 0$$

Generally speaking, suppose F_j and W_i are the opportunity sets.

$$\left| JP(i \in F_j | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{K_F \exp\{\tilde{U}_{ij}\}}{1 + \frac{1}{J}\sum_{k=1}^{J}\exp\{\tilde{U}_{ik}\}} \right| \to 0,$$

$$\left| P(i \notin F_k, \forall k = 1, ..., J | \tilde{U}_{i1}, ..., \tilde{U}_{iJ}) - \frac{1}{\left(1 + \frac{1}{J}\sum_{k=1}^{J}\exp\{\tilde{U}_{ik}\}\right)^{K_F}} \right| \to 0.$$

Similarly,

$$\left| JP(j \in W_i|\tilde{V}_{j1}, ..., \tilde{V}_{jJ}) - \frac{K_W \exp\{\tilde{V}_{ji}\}}{1 + \frac{1}{J}\sum_{k=1}^{J} \exp\{\tilde{V}_{jk}\}} \right| \to 0,$$

$$\left| P(j \notin W_k, \forall k = 1, ..., J|\tilde{V}_{j1}, ..., \tilde{V}_{jJ}) - \frac{1}{\left(1 + \frac{1}{J}\sum_{k=1}^{J} \exp\{\tilde{V}_{jk}\}\right)^{K_W}} \right| \to 0.$$

□

Proof of Proposition 1 . Define the number of workers available to firm i

$$J_{F_i} = \sum_{j=1}^{n_W} \mathbb{1}\{V_{ji} \geq V_j(F_j)\},$$

the number of firms available to worker j

$$J_{W_j} = \sum_{i=1}^{n_F} \mathbb{1}\left\{U_{ij} \geq U_i(W_i)\right\}.$$

Then

$$J_{W_j} = \sum_{i=1}^{n_F} \mathbb{1}\{U_{ij} \geq U_i(W_i)\} \leq \sum_{i=1}^{n_F} \mathbb{1}\{U_{ij} \geq U_{iO_{(K_F)}}\} = \sum_{i=1}^{n_F} \mathbb{1}\{i \in \bar{F}_j\} := \bar{J}_{W_j}$$

Since

$$P\left(U_{ij} \geq U_{iO_{(K_F)}} \middle| x_i, z_j\right) = 1 - P\left(U_{iO_{(K_F)}} > U_{ij} \middle| x_i, z_j\right) \to 1 - \frac{1}{\left(1 + \frac{1}{J}\exp\{\tilde{U}_{ij}\}\right)^{K_F}},$$

$$E[\bar{J}_{W_j}|z_j, x_1, ..., x_{n_F}] \to \sum_{i=1}^{n_F} \left(1 - \frac{1}{\left(1 + \frac{1}{J}\exp\{\tilde{U}_{ij}\}\right)^{K_F}}\right) \leq n_F \left(1 - \left(1 + \frac{1}{J}\exp\{\bar{U}\}\right)^{-K_F}\right)$$

$$\overset{J\to\infty}{=} n_F \left(1 - \exp\left\{-\frac{K_F}{J}\exp\{\bar{U}\}\right\}\right).$$

J should grow at the same rate as the size of the opportunity set J_{W_j}.

If $K_F/J \to 0$,

$$
\begin{aligned}
& E[\bar{J}_{W_j}|z_j, x_1, ..., x_{n_F}] \leq \frac{n_F K_F}{J} \exp\{\bar{U}\} \\
& \Rightarrow \quad J = O(\sqrt{n}),\ E[\bar{J}_{W_j}|z_j, x_1, ..., x_{n_F}] = O(\sqrt{n}),\ \alpha > 1/2.
\end{aligned}
$$

If $K_F/J = O(1)$,

$$
\begin{aligned}
& E[\bar{J}_{W_j}|z_j, x_1, ..., x_{n_F}] = O(n_F) = O(n^\alpha),\ J = O(n^\alpha),\ K_F = O(n^\alpha) \\
& \Rightarrow \quad \alpha = 1/2,\ J = O(\sqrt{n}), E[\bar{J}_{W_j}|z_j, x_1, ..., x_{n_F}] = O(\sqrt{n}).
\end{aligned}
$$

If $K_F/J = O(n^\beta)$, contradiction as K_F is not restricting, and thus K_W not restricting. Should be similar to many-to-one matching.

$$
\begin{aligned}
& E[\bar{J}_{W_j}|z_j, x_1, ..., x_{n_F}] = O(n_F) = O(n^\alpha),\ J = O(n^\alpha),\ K_F = O(n^{\alpha+\beta}) \\
& \Rightarrow \quad \alpha < 1/2,\ \beta = 1 - 2\alpha,\ J = O(\sqrt{n}), E[\bar{J}_{W_j}|z_j, x_1, ..., x_{n_F}] = O(\sqrt{n}).
\end{aligned}
$$

For the following I focus on $\alpha > 1/2$. ($\alpha < 1/2$ capacity constraint not binding when $n \to \infty$)

By law of iterated expectation,

$$
E[\bar{J}_{W_j}] \leq \sqrt{n} \left(\exp\{\bar{U} + \gamma_F + \gamma'_F\} + o(1) \right).
$$

Rate of Variance of $\bar{J}_{W_j}$: Since

$$
\frac{K_F}{J} \exp\{-\tilde{U}\} \leq p_{ijn} := 1 - \frac{1}{\left(1 + \frac{1}{J}\exp\{\tilde{U}_{ij}\}\right)^{K_F}} \leq \frac{K_F}{J} \exp\{\tilde{U}\}
$$

$$
\bar{v}_{jn} = \frac{1}{n_F} \sum_{i=1}^{n_F} p_{ijn}(1 - p_{ijn}) = O\left(\frac{K_F}{J}\right) = O\left(n^{1/2-\alpha}\right)
$$

Apply a CLT for independent heterogeneously distributed random variables,

$$\frac{\frac{1}{n_F}\left(\bar{J}_{W_j}-E\left[\bar{J}_{W_j}\right]\right)}{\sqrt{\bar{v}_{jn}/n_F}}=\frac{1}{\sqrt{\bar{v}_{jn}n_F}}\sum_{i=1}^{n_F}\left(\mathbb{1}\left\{U_{ij}\geq U_{iO(K_F)}\right\}-p_{ijn}\right)\xrightarrow{d}\mathcal{N}(0,1)$$

$$\frac{\bar{J}_{W_j}-E\left[\bar{J}_{W_j}\right]}{O(n^{1/4})}\xrightarrow{d}\mathcal{N}(0,1)$$

We obtain

$$\frac{\bar{J}_{W_j}-E\left[\bar{J}_{W_j}\right]}{o(n^{1/2})}\xrightarrow{p}\mathcal{N}(0,1).$$

Similarly, we can show that for $\bar{J}_{F_i}$. and the lower bounds.

Lower bound $J^{\circ}_{W_i}$: Denote the set of worker j that prefer firm i to their outside option or any firm in $\bar{F}_j$ by W°_i.

$$J_{W_j}=\sum_{i=1}^{n_F}\mathbb{1}\left\{U_{ij}\geq U_i(W_i)\right\}\geq\sum_{i=1}^{n_F}\mathbb{1}\left\{U_{ij}\geq U_i(\bar{W}_i)\right\}=\sum_{i=1}^{n_F}\mathbb{1}\left\{i\in F^{\circ}_j\right\}:=J^{\circ}_{W_j}$$

Similar for $J^{\circ}_{F_i}$.

In sum, when $\alpha>1/2$, sizes of opportunity sets F_j and W_i are $O(\sqrt{n})$.

With a little more effort we can show the rate of growth of the size of the opportunity set is $\sqrt{n}$. What is the rate of growth of J we should choose? Same rate as the rate of growth of the size of opportunity set. Alpha is the growth rate of number of firms. When α is less than 1/2, as n approaches infinity, the capacity constraint will not be binding since the number of firms does not grow fast enough. (Second, we demonstrate that dependence of taste shifters and opportunity sets is negligible for CCPs when n is large. Hence we can approximate choice probabilities using the inclusive values.)

In sum, when $\alpha>1/2$, sizes of opportunity sets F_j and W_i are $O(\sqrt{n})$.

$\square$

Proof of Theorem 1. Suppose that $J=\sqrt{n}$.

Let

$$I_{f_i} = I_{f_i}[W_i] = \frac{1}{\sqrt{n}} \sum_{k \in W_i} \exp\{\tilde{U}_{ik}\}$$
$$I_{w_j} = I_{w_j}[F_j] = \frac{1}{\sqrt{n}} \sum_{k \in F_j} \exp\{\tilde{V}_{jk}\}$$

Then

$$\sqrt{n} P(i \in F_j | x_i, z_j, I_{f_i}) = \frac{K_F \exp\{U(x_i, z_j)\}}{1 + I_{f_i}} + o_p(1)$$
$$P(i \notin F_k, \forall k \in W_i | x_i, z_j, I_{f_i}) = \frac{1}{(1 + I_{f_i})^{K_F}} + o_p(1)$$

Similarly,

$$\sqrt{n} P(j \in W_i | z_j, x_i, I_{w_j}) = \frac{K_W \exp\{V(z_j, x_i)\}}{1 + I_{w_j}} + o_p(1)$$
$$P(j \notin W_k, \forall k \in F_j | z_j, x_i, I_{w_j}) = \frac{1}{\left(1 + I_{w_j}\right)^{K_W}} + o_p(1)$$

Expected inclusive value function

$$\hat{\Gamma}_f(x_i) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n_W} \exp U(x_i, z_k) P(k \in W_i | z_k, x_i, I_{w_k}) = \frac{K_W}{n} \sum_{k=1}^{n_W} \frac{\exp\{U(x_i, z_k) + V(z_k, x_i)\}}{1 + I_{w_k}}$$
$$\hat{\Gamma}_w(z_j) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n_F} \exp V(z_j, x_k) P(k \in F_j | z_j, x_k, I_{f_k}) = \frac{K_F}{n} \sum_{k=1}^{n_F} \frac{\exp\{V(z_j, x_k) + U(x_k, z_j)\}}{1 + I_{f_k}}$$

Or rewrite it using a fixed-point characterization,

$$\hat{\Gamma}_f(x) = \frac{K_W}{n} \sum_{k=1}^{n_W} \frac{\exp\{U(x, z_k) + V(z_k, x)\}}{1 + \hat{\Gamma}_w(z_k)}$$
$$\hat{\Gamma}_w(z) = \frac{K_F}{n} \sum_{k=1}^{n_F} \frac{\exp\{V(z, x_k) + U(x_k, z)\}}{1 + \hat{\Gamma}_f(x_k)}$$

□

Proof of Lemma 2. Note that $T = \sigma^{-1}\left(U_{im_{(K_F+1)}} - \tilde{U}_{ij}\right) = \eta_{im_{(K_F+1)}} + \sigma^{-1}\left(\tilde{U}_{im_{(K_F+1)}} - \tilde{U}_{ij}\right)$

$$\begin{aligned}
F_T(t) &= Pr\left(T \leq t | \tilde{U}_{im_{(K_F+1)}}, \tilde{U}_{ij}\right) \\
&= Pr\left(\eta_{im_{(K_F+1)}} \leq t + \sigma^{-1}\left(\tilde{U}_{ij} - \tilde{U}_{im_{(K_F+1)}}\right) | \tilde{U}_{im_{(K_F+1)}}, \tilde{U}_{ij}\right) \\
&= \prod_{\substack{m \neq m_{(i)} \\ i=1,\ldots,K_F}} \Phi\left(t + \sigma^{-1}\left(\tilde{U}_{ij} - \tilde{U}_{im}\right)\right) \\
&= \exp\left\{\frac{1}{J}\sum_{m=1}^{2J} J \log \Phi\left(t + \sigma^{-1}\left(\tilde{U}_{ij} - \tilde{U}_{im}\right)\right)\right\} \Big/ \left\{\prod_{\substack{m = m_{(i)} \\ i=1,\ldots,K_F}} \Phi\left(t + \sigma^{-1}\left(\tilde{U}_{ij} - \tilde{U}_{im}\right)\right)\right\}
\end{aligned}$$

Let $b_J := \Phi^{-1}(1 - \frac{1}{J}) \to +\infty$, $a_J = a(b_J) = \sigma^{-1}$, where $a(t) = \frac{1-\Phi(t)}{\phi(t)}$. Let

$$R_J(t) = \frac{1}{J}\sum_{m=1}^{2J} J \log \Phi\left(t + \sigma^{-1}\left(\tilde{U}_{ij} - \tilde{U}_{im}\right)\right).$$

By a change of variable, $t = a_J s + b_J$,

$$R_J(s) = \frac{1}{J}\sum_{m=1}^{2J} J \log \Phi\left(b_J + \sigma^{-1}\left(s + \tilde{U}_{ij} - \tilde{U}_{im}\right)\right)$$

Since $-\log G \approx 1 - G$,

$$R_J(s) = -\frac{1}{J}\sum_{m=1}^{2J} J\left\{1 - \Phi\left(b_J + a_J\left(s + \tilde{U}_{ij} - \tilde{U}_{im}\right)\right)\right\} + o(1).$$

And with $\frac{1}{J} = 1 - \Phi(b_J)$, we have

$$\begin{aligned}
R_J(s) &= -\frac{1}{J}\sum_{m=1}^{2J} J\left\{1 - \Phi\left(b_J + a_J\left(s + \tilde{U}_{ij} - \tilde{U}_{im}\right)\right)\right\} + o(1) \\
&= -\frac{1}{J}\sum_{m=1}^{2J} \frac{1 - \Phi\left(b_J + a_J\left(s + \tilde{U}_{ij} - \tilde{U}_{im}\right)\right)}{1 - \Phi(b_J)} + o(1) \\
&= -e^{-s} \cdot \frac{1}{J}\sum_{m=1}^{2J} \exp\{\tilde{U}_{im} - \tilde{U}_{ij}\} + o(1)
\end{aligned}$$

since $J(1-\Phi(b_J+a_J s)) = \frac{1-\Phi(b_J+a_J s)}{1-\Phi(b_J)} \to e^{-s}$. In addition,

$$\begin{aligned}
\prod_{\substack{m=m_{(i)}\\ i=1,\ldots,K_F}} \Phi\left(t+\sigma^{-1}\left(\tilde{U}_{ij}-\tilde{U}_{im}\right)\right) =& \prod_{\substack{m=m_{(i)}\\ i=1,\ldots,K_F}} \Phi\left(b_J+a_J\left(s+\tilde{U}_{ij}-\tilde{U}_{im}\right)\right) \\
=& \prod_{\substack{m=m_{(i)}\\ i=1,\ldots,K_F}} (1-J^{-1}\exp\{-(s+\tilde{U}_{ij}-\tilde{U}_{im})\}) \\
\to& 1
\end{aligned}$$

Thus

$$|F_T(t)-\exp\{-e^{-s}\cdot C\}| = |F_T(t)-\exp\{-\exp\left\{-\frac{t-\Phi^{-1}(1-1/J)}{\sigma^{-1}}+\log C\right\}\}| \to 0$$

where $C = \frac{1}{J}\sum_{m=1}^{2J}\exp\{\tilde{U}_{im}-\tilde{U}_{ij}\}$. Therefore,

$$F_T(t) \to Gumbel(t; \Phi^{-1}(1-1/J)+\sigma^{-1}\log C, \sigma^{-1}).$$

□

Proof of Theorem 2 .

$$\begin{aligned}
E[\sigma\eta_{ij}|X,Z,A_{ij}=1] =& E\left[\sigma\eta_{ij}|X,Z,U^*_{ij}\geq U^*_{im_{(K_F)}}, V^*_{ji}\geq V^*_{jn_{(K_W)}}\right] \\
=& E\left[\sigma\eta_{ij}|\sigma\eta_{ij}\geq\sigma\eta_{im_{(K_F+1)}}+\left(\tilde{U}_{im_{(K_F+1)}}-\tilde{U}_{ij}\right)\right]
\end{aligned}$$

Let $T=\sigma^{-1}\left(U_{im_{(K_F+1)}}-\tilde{U}_{ij}\right)=\eta_{im_{(K_F+1)}}+\sigma^{-1}\left(\tilde{U}_{im_{(K_F+1)}}-\tilde{U}_{ij}\right)$.

$$E[\sigma\eta_{ij}|X,Z,A_{ij}=1] = E\left[\sigma\eta_{ij}|\eta_{ij}\geq T\right] = \sigma E_T\left[E\left[\eta_{ij}|\eta_{ij}\geq T, T=t\right]\right] = \sigma E_T\left[\frac{\phi(t)}{1-\Phi(t)}\right]$$

Note that $T\sim Gumbel(t;\Phi^{-1}(1-1/J)+\sigma^{-1}\log C,\sigma^{-1})$. As $J\to\infty$, the Gumbel distribution is skewed to the right. Since $\frac{\phi(t)}{1-\Phi(t)}\to t$ when $t\to\infty$,

$$E_T\left[\frac{\phi(t)}{1-\Phi(t)}\right] \to E_T(t).$$

For $T \sim Gumbel(t; \mu, \beta)$, $E_T(t) = \mu + \beta\kappa$, where $\kappa \approx 0.5772$ is the Euler's constant. Thus

$$E[U_{ij}|X, Z, A_{ij} = 1] = \tilde{U}_{ij} + \log C + 1 + \kappa.$$

□

1.7 B. Technical Definitions

We start with the definition of *pairwise stability.*

Definition 1 (Pairwise Stability)**.** *A feasible matching μ is pairwise-stable, if there are no firm and worker who are not partners, but can both obtain a preferred set of partners by becoming partners, while possibly dissolving other partnerships of μ to remain within their quotas and keeping other ones.*

That is, μ must satisfy the conditions

1. *if $U_{ij} > U_{i(K_F)}$, $j \notin \bar{\mu}_F(i)$, then $V_{ji} \leq V_{j(K_W)}$;*
2. *if $V_{ji} > V_{j(K_W)}$, $i \notin \bar{\mu}_W(j)$, then $U_{ij} \leq U_{i(K_F)}$.*

As an extension, Roth (1984)[48] and Sotomayor (1999)[55] proposed *setwise stability.*

Definition 2 (Setwise Stability)**.** *A matching μ will be called setwise-stable if there is no subset of agents who by forming new partnerships only among themselves, possibly dissolving some partnerships of μ to remain within their quotas and possibly keeping other ones, can all obtain a strictly preferred set of partners.*

Echenique and Oviedo (2006)[19] gives conditions (*substitutability* and *strong substitutability*) under which the setwise-stable set is nonempty and can be approached through an algorithm called *T-algorithm.* Define the *fix-point set* by the matchings where each agent a is choosing her best set of partners, out of the set of potential partners who are willing to link to a given their current match. And define the set of *pre-matchings* by $\mathcal{V} = \mathcal{V}_F \times \mathcal{V}_W$, where

$\mathcal{V}_F = (2^W)^F$, $\mathcal{V}_W = (2^F)^W$ since $\nu_F : F \to 2^W$, $\nu_W : W \to 2^F$. We often refer to $\nu_W(w)$ by $\nu(w)$ and to $\nu_F(f)$ by $\nu(f)$.

In the firm-worker setup, we define the *map* T as follows.

Definition 3 (Map T). *Let ν be a pre-matching, and let*

$$W(f, \nu) = \{w \in W : f \in Ch\,(\nu(w) \cup \{f\}, P(w))\}$$

and

$$F(w, \nu) = \{f \in F : w \in Ch\,(\nu(f) \cup \{w\}, P(f))\},$$

where the set $F(w, \nu)$ is the set of firms f that are willing to hire w, possibly after firing some of the workers it was assigned by ν. The set $W(f, \nu)$ is the set of workers w that are willing to add f to its set of firms $\nu(w)$, possibly after firing some firms in $\nu(w)$.

Now, define $T : \mathcal{V} \to \mathcal{V}$ by

$$(T\nu)(s) = \begin{cases} Ch(W(s, \nu), P(s)), & \text{if } s \in F \\ Ch(F(s, \nu), P(s)), & \text{if } s \in W \end{cases}$$

We interpret the map T using the firm-worker setup. $(T\nu)(f)$ is firm f's optimal team of workers, among those willing to work for f, and $(T\nu)(w)$ is the set of firms preferred by w, among the firms that are willing to hire w.

Let the *fix-point set* be the set of fixed points of T; we denote it by $\mathcal{E}(P)$. Then $\mathcal{E}(P) = \{\nu \in \mathcal{V} : \nu = T\nu\}$.

We now describe an algorithm that is associated with the techniques we use to prove our results: the techniques exploit the fixed points of T, and the algorithm is designed to find a fixed point of T. The definition of the algorithm is as follows.

Definition 4. *(T-algorithm) The T-algorithm is the procedure of iterating T, starting at some pre-matching ν.*

Let ν_0 and ν_1 be the pre-matchings defined by $\nu_0(f) = \nu_1(w) = \emptyset$, $\nu_0(w) = F$, and

$\nu_1(f) = W$ for all w and f. We consider the T-algorithm starting at pre-matchings ν_0 and ν_1. Then Echenique and Oviedo (2006)[19] propose the following theorem.

Theorem 3 (Fixed-point Set)**.** *If $P(F)$ is substitutable and $P(W)$ is strongly substitutable, then*

(i) $\mathcal{E}(P) = SW(P)$

(ii) $SW(P)$ *is nonempty. The T-algorithm finds a matching in* $SW(P)$*,*

where $\mathcal{E}(P)$ *is the fixed-point set that can be approached through the T-algorithm.*

We consider two restrictions on agents' preferences. The first is substitutability, first introduced by Kelso and Crawford (1982) and used extensively in the matching literature. The second is a strengthening of substitutability that we call strong substitutability.

Definition 5. *(Substitutability) An agent a's preference ordering $P(a)$ satisfies substitutability if, for any sets S and S', with $S \subseteq S'$,*

$$b \in Ch(S' \cup b, P(a)) \textit{ implies } b \in Ch(S \cup b, P(a)).$$

Say that a preference profile P is substitutable if $P(a)$ satisfies substitutability for every agent a.

Definition 6. *(Strong Substitutability) An agent a's preference ordering $P(a)$ satisfies strong substitutability if, for any sets S and S', with $S'\ P(a)\ S$,*

$$b \in Ch(S' \cup b, P(a)) \textit{ implies } b \in Ch(S \cup b, P(a)).$$

Say that a preference profile P is strongly substitutable if $P(a)$ satisfies strong substitutability for every agent a.

We explain the difference of the two definitions. Let f be a firm. Substitutability of firm f's preferences requires that if hiring w is optimal when the set of available workers is $\{w\} \cup S'$, and S is a subset of S', then hiring w must still be optimal when the set of available workers is $\{w\} \cup S$. Or equivalently, if w is chosen from a given set of workers, she is chosen also from a smaller set of workers. Strong substitutability requires that if hiring w is optimal when the set of available workers is $\{w\} \cup S'$, and the firm prefers S' to S, then hiring w must still be optimal when the set of available workers is $\{w\} \cup S'$. Or equivalently, if w is chosen from a given set of workers, she is chosen also from a worse set of workers.

CHAPTER 2

The Estimation of Diffusion Processes with Private Network Information

2.1 Introduction

Diffusion processes, in which an innovation spreads from one person to another and eventually reaches a substantial population, are prevalent phenomena in social or geographic networks. Evidence shows that networks play a significant role in information diffusion.[1] In particular, policymakers and businesses often rely on social networks to diffuse information to the community. As one important example, the spread of Gmail started from a small group of initial users and cascaded to affect a significant proportion of the population via chains of invitations. Other empirical evidence includes the spread of new products and services through viral marketing campaigns and referral programs (Leskovec et al., 2007[38]), the uptake of microfinance in developing countries (Banerjee et al., 2013[6]), and the spread of rumors about India's demonetization policy (Banerjee et al., 2018[5]). Therefore, a model for diffusion processes in social networks is crucial for designing marketing strategies that maximize diffusion and launching new policies that enhance social welfare.

This paper provides a structural econometric analysis of diffusion processes in social networks. In the diffusion process, a message about a new product or a new idea spreads through social ties. Once exposed by their neighbors, agents decide whether to adopt the new product and pass on the information. We study a diffusion model with players who observe their neighbors and make decisions based on their beliefs about their neighbors' decisions. We are interested in the interpersonal influences governing agents' information and the decisions that agents make. Extending Sadler (2020)[51], we build a tractable model for the structural analysis of diffusion processes. We show that a unique, symmetric equilibrium exists under certain conditions and propose a consistent two-step estimation approach for individual payoffs using only a single large network where the number of players approaches infinity.

[1] An old and large empirical literature studies the role of social networks in information diffusion. Ryan and Gross (1943)[50] and Coleman, Katz, and Menzel (1966)[14] highlight the importance of social connections in technology adoption. More recent literature emphasizes the players' strategic interactions in the network, for example, social learning in agriculture that influences the uptake of new technologies (Griliches 1957[28], Foster and Rosenzweig 1995[21], Munshi 2004[45], Bandiera and Rasul 2006[3], Conley and Udry 2010[15], Burlig and Stephens 2019[10]), and the impact of seed players on diffusion results (Banerjee et al. 2013[6]).

We illustrate diffusion processes using the example from Banerjee et al. (2013)[6] that depicts the diffusion of a new microfinance program through word-of-mouth. Suppose all players are connected by a social network graph. Initially, none of the players in this population has been exposed to the product (e.g., the opportunity to obtain microfinance loans). Only a seed player is informed in the first period. When a player becomes aware, she makes an irreversible decision whether to adopt the action, that is, to apply for the loan. Once a player adopts, she informs her neighbors about the microfinance lender, and thus her neighbors become aware of the opportunity. This step is repeated until there are no new adopters. A player's payoff from adoption is based on the decisions of her neighbors, for example, the number of neighbors who will participate eventually, because they benefit from information-sharing. Thus, once a player is aware, she forms beliefs about her neighbors' strategies and makes the adoption decision based on her expected payoffs. This strategic interaction among agents leads to correlated individual decisions, which is key to estimating social effects.

The main question is how to estimate the payoff parameters if we have data on the diffusion process, including the network structure, the information transmission path, and the participation decisions. The payoff parameters are useful for welfare analysis and out-of-sample forecasting.

We model diffusion processes in fixed social networks where players only observe their close neighbors. The theoretical literature focuses on two main modeling approaches: mean-field modeling and a fixed network approach. The first approach relies on mean-field approximations[2] and is frequently applied in empirical research. The classic Bass diffusion model (Bass 1969[8]) assumes that homogeneous adopters affect all other players equally. More recent work studies the impact of network structure on diffusion processes, allowing nodes with different degrees (Jackson and Rogers, 2007[32], López-Pintado, 2006[40], 2008[41], 2012[42]) and types (Jackson and López-Pintado, 2013[31]). Moreover, the estimation of mean-

[2]Mean-field approximations assume (a) the speed of diffusion at any instant depends only on the fraction of adopters in the population; (b) the population is a continuum, so individual adoption decisions are insignificant to final diffusion results.

field models has been extensively explored in the econometrics literature.[3] The limitation of these mean-field models is obvious: they are analogous to models where the links shuffle and reconnect every period, that is, neighbors are randomly assigned and resampled every period. Thus, the strategic effects of individual players are minimized, and the diffusion outcome is determined by the network's degree distribution.

Alternatively, more recent papers, including our paper, focus on diffusion processes on fixed networks where one player's adoption decision could possibly affect the diffusion outcome. In a fixed network where the links are persistent, individual decisions depend on their permanent neighbors, and players react strategically to their neighbors' adoption decisions. Morris (2000)[44] and Watts (2002)[61] study myopic adoption strategies based on a threshold rule such that players adopt a behavior if enough neighbors did so in the previous period. The networks discussed in these two papers assume homogeneous players, with only a single type of nodes and edges. In contrast, Sadler (2020)[51] proposes a model that incorporates Bayesian players that are heterogeneous in types, that is, a more complicated network consisting of heterogeneous nodes and edges where the agents in this network game make strategic choices. Despite the empirical relevance and theoretical importance, diffusion processes on fixed networks have not been investigated in the structural econometrics literature. The present paper follows Sadler (2020)[51] to study heterogeneous strategic players and provides an econometric framework for the structural analysis of diffusion processes. Unlike Sadler (2020)[51], where neighbor types remain unobservable to players, our paper assumes players observe the neighbor types because people often have some information about their friends' preferences when making decisions.

The contribution is mainly twofold. First, this paper provides an econometric framework for analyzing diffusion processes in fixed networks with strategic players that observe their close neighbors. We extend Sadler (2020)[51] to build a model of network games where

[3] Schmittlein and Mahajan (1982)[53] propose a maximum likelihood estimator (MLE) for Bass model parameter estimation and apply it to model forecasting. Srinivasan and Mason (1986)[56] propose a nonlinear least squares (NLS) approach that considers not only sampling errors but also other errors. They report that the NLS is comparable to the MLE in terms of fit and predictive validity. Boswijk and Franses (2005)[9] develop an asymptotic theory for the parameters in an alternative version of the Bass model that incorporates heteroscedastic errors.

the individuals are heterogeneous in types and hold private information about their links and payoffs. Instead of the assumption that players do not observe neighbor types, as in Sadler (2020)[51], we assume that neighbor information is observable to players. Players learn about an innovation from their neighbors and make adoption decisions that maximize their expected payoffs, given beliefs about their neighbors' propensity to adopt. A key feature of the model, which we prove in this paper, is the existence of equilibrium beliefs formed by strategic players. In addition, since a fundamental difficulty of this procedure is that diffusion models generally have multiple equilibria for beliefs (Jackson and Yariv, 2007[34]), we provide conditions that guarantee the unique Bayesian equilibrium of the model. We show that the beliefs are formed based on anonymized information such as types and neighborhood characteristics. We characterize unique, symmetric equilibrium beliefs.

Furthermore, this paper develops a feasible two-step procedure to identify and estimate individual payoffs with only a single large network. We first estimate the equilibrium beliefs via sample analogs and then the payoff parameters using a maximum likelihood estimator. This two-step estimator for payoff parameters is consistent under regularity conditions and tractable with observations from a single large network. By restricting the range of payoff-relevant connections and assuming symmetric equilibria – only neighbors with direct links can affect one's payoff, and agents form the same belief about observationally equivalent neighbors – we deal with the computational difficulties that arise from the highly heterogeneous strategic environments across agents. These assumptions enable the estimation of equilibrium beliefs. In addition, because observations are history-dependent in diffusion processes – that is, players make decisions only if they are informed – we address the possibly selective observations by showing that the forward type and degree distribution of the newly informed group in each period is independent of the payoff parameters of interest.

Our model is closely related to the literature that investigates the social interaction models where individuals only observe their close neighbors.[4] Unlike the standard literature that assumes players observe the whole network structure, network games with such incomplete

[4]This assumption on incomplete information is prevalent when the size of the network is large (Banerjee et al., 2014[7]).

information often adopt a Bayesian Nash equilibrium as their equilibrium concept. Galeotti et al.(2010)[23] develop a general framework for static incomplete information games and establish the existence of symmetric, monotone Bayesian Nash equilibria. Jackson and Yariv (2007)[34] establish the correspondence between the Bayesian equilibria in a static incomplete information game and the steady state of a mean-field diffusion model. While these theory papers study Bayesian equilibria in static games or mean-field processes, we demonstrate the existence of a Bayesian equilibrium in diffusion games with fixed networks.

A recent strand of econometrics literature tackles the challenge that arises from incomplete network information. Canen et al. (2020)[12] take a behavioral approach to modeling networks games and provide an estimation method for a model of linear local interactions. Their behavioral approach assumes that a player projects his or her own beliefs about others' strategies onto his neighbors' beliefs, which may be restrictive in application. In contrast, I demonstrate a symmetric equilibrium where a player's equilibrium beliefs only depend on anonymized information such as the type and the degree of the player and her neighbors' types. Eraslan and Tang (2017)[20] characterize a unique, symmetric Bayesian Nash equilibrium and propose a two-step m-estimator for individual payoffs in the simultaneous-move setup. Their paper characterizes the equilibrium by the conditional expectation of players' actions given their types and degrees. Thus, a player's expected payoff, given that neighbors' degrees are unobservable, depends on the conditional expectations of all possible degrees. We instead characterize the equilibrium by players' beliefs about their neighbors' strategies given players' types and degrees and neighbors' types. Therefore, the estimation is considerably simplified because we do not need to integrate neighbors' degrees. An additional difference from their paper is that our paper applies to diffusion games, so that the observations are potentially correlated, because players are informed only when someone in their neighborhood decides to adopt.

The literature on diffusion processes also lies at the intersection of marketing and computer science. The diffusion of innovation has been a major interest in the marketing literature. Abrahamson and Rosenkopf (1997)[1] study the social network effects using computer simulations. They show that the extent of diffusion is affected by the structure of social networks because the information is channeled only to certain potential adopters. Goldenberg et al. (2009)[27] and Iyengar et al. (2011)[30] discuss the roles of central individuals in diffusion rate and extent. By incorporating network structure into the diffusion model, the literature is extending its focus from a homogeneous and fully connected social system to a more complicated and possibly heterogeneous network (see Peres et al. (2007)[47] for an overview).

The model for diffusion processes in fixed networks exhibits significant potential for future research. First, it provides a tool for out-of-sample forecasting. We can estimate the payoff parameters using data obtained from other communities and perform predictions over the network of interest. Second, with a well-defined preference structure, the model helps determine the group of early adopters (or the seeds) that maximizes the speed and spread of product adoption and brings the best referral promotion outcomes.[5] Lastly, the model allows us to relate the estimated individual payoff parameters with a series of player-specific characteristics.[6]

This paper is organized as follows. Sections 2.2 lays out the model of diffusion processes, establishes the existence of equilibria, and characterizes the unique, symmetric equilibrium under certain conditions. Section 2.3 demonstrates the identification strategy as the number of individuals approaches infinity while observing a single network. In Section 2.4, we present a two-step estimator for individual payoffs. Section 2.7 provides Monte Carlo simulation results, and Section 2.8 concludes. In Appendix 2.9, we discuss the assumptions of the model. Proofs are collected in Appendix 2.10.

[5]In various applications (Van der Lans et al., 2010[60]; Banerjee et al., 2013[6]; Cai et al., 2015[11]), early adopters are thought to affect the diffusion outcomes.

[6]For mean-field models, there is extensive marketing literature that relates the diffusion parameters with country-specific characteristics, and thus allows multinational diffusion processes (Gatignon et al., 1989[25]; Takada and Jain, 1991[58]; Helsen et al., 1993[29]; Desiraju et al., 2004[16]; Talukdar et al., 2002[59]). We can perform the same study on discrete networks.

2.2 The Diffusion Model

To illustrate the data observed from a diffusion process, we use a referral program as an example. Suppose that each time a person purchases a product, she will send referral emails to her friends. For each consumer i who is informed of this product in the recommendation network, a researcher observes i's adoption decision, the person who send the email to i, i's type, and her neighbors' types. The recommendation network is pre-determined and exogenously given. It describes the long-term relationship among all consumers and can be recovered from all historical recommendation records on all products.

The following model characterizes the data generating process.

Players. We focus on the sequence of games such that the number of players n approaches infinity. For each n-th game, let $\mathcal{N} = \{1, ..., n\}$ be the set of n agents. At the beginning of the game, each agent $i \in \mathcal{N}$ is endowed with a type $x_i \in \mathbf{X}$, where $\mathbf{X} = \{X_1, ..., X_K\}$ is a discrete finite support, and an idiosyncratic shock $\varepsilon_i \in \mathbb{R}$.

Network. We consider an exogenous undirected network where all players have no more than M links. Denote the network by an n-by-n matrix $G \triangleq (G_{ij})_{i,j\in\mathcal{N}}$. $G_{ij} = G_{ji} = 1$ if agent i is connected to agent j and $G_{ij} = 0$ if otherwise. Denote the set of i's neighbors by $N_i \equiv \{j \in \mathcal{N} : G_{ij} = 1\}$, i.e., the agents directly connect to i. Then denote the neighborhood attribute by $\mathbf{d}_i$. $\mathbf{d}_i = (d_{i1}, d_{i2}, ..., d_{iK})$ is a vector of size K, where each d_{ik} denotes the number of type-X_k neighbors that agent i has. We call $\mathbf{d}_i$ the degree or the neighborhood attribute of agent i in the rest of the paper.

There exists a *player type distribution* $T \in \Delta(\mathbf{X})$ and a *degree distribution* $D_{\mathbf{X}} \in \Delta(\mathbb{N}^K)$ for each type with maximum total degrees M, i.e., the total number of neighbors of all types should not exceed M. The player type and degree distributions are common knowledge.

Consider the asymptotics that $n \to \infty$. For each $n \in \mathbb{N}$, we generate a network with type and degree sequences $\left(\mathbf{x}^{(n)}, \mathbf{d}^{(n)}\right)$. Let $N_{x,\mathbf{d}}^{(n)}$ denote the number of nodes with type x

and degree $\mathbf{d}$ in the n-th game. Assume for each type $x \in \mathbf{X}$ and degree $\mathbf{d} \in \mathbb{N}^K$,

$$\lim_{n\to\infty} \frac{N_{x,\mathbf{d}}^{(n)}}{n} = \mathbf{P}(T = x) \cdot \mathbf{P}(D_{\mathbf{X}} = \mathbf{d}),$$

the network converges in distribution as n approaches infinity.

Diffusion process. The game characterizes the diffusion of a behavior on a network of the n players over n periods $t = 0, 1, ..., n-1$[7]. Denote individual i's action at time t by $y_i(t) \in \mathcal{Y} = \{0, 1\}$, that is, $y_i(t) = 1$ if agent i adopts the new product at time t, $y_i = 0$ if agent i does not. Initially $y_i(0) = 0$ for all players, and none of the agents has been informed of the behavior. At time 0, nature does three things: first, it draws a type and degree sequence $\{(\mathbf{x}^{(n)}, \mathbf{d}^{(n)})\}_{n\in\mathbb{N}}$ from the distribution $(T, D_{\mathbf{X}})$ and forms a random network; second, it assigns the idiosyncratic shocks to the n players in the network; third, it draws a seed player uniformly at random from all players to adopt the behavior and pass the information to all her neighbors.

In each following period, those informed make irreversible decisions whether to adopt. Once a player adopts, her neighbors become informed. Players have to decide whether to adopt when they become informed, and individuals who refrain from adopting cannot revisit their decision. Player i acts at time $t_i = 1 + \min\{t : \exists j \in N_i \text{ such that } y_j(t) = 1\}$. Let $y_i = y_i(n-1)$ denote whether player i has adopted the behavior in this process.

Parent(s). We use ω_i to denote agent i's parent(s), meaning the neighbor(s) who already adopted and passed the information to agent i before agent i makes her decision. In our setup, the probability that an agent has more than one parent is negligible as n becomes large. Sadler (2020)[51] shows that, as n grows, the local network structure of a configuration

[7]In a network with n players, the maximum number of periods needed to complete the diffusion process is n.

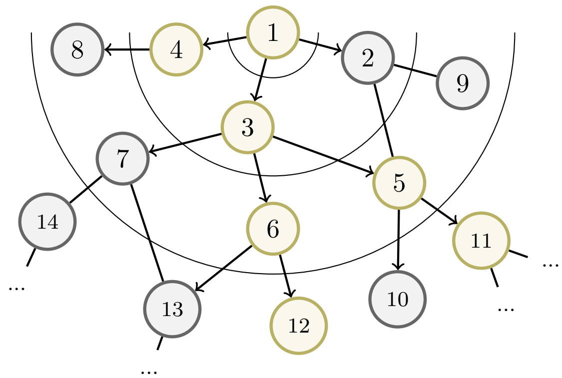


Figure 2.1: A diffusion process. The olive nodes are players that adopt. Arrows indicate the transition of information. Each arrow starts from a referrer and ends in a receiver.

model[8] converges in distribution to a branching process[9] as the chance of finding a cycle approaches zero. In simulations, we find that, in a network with 1,000 players, fewer than 2% of the players have more than one parent. Because we focus on large networks with a substantial proportion of adopters for estimation purposes, we can safely assume that each player has only one parent.

Information. Let $\tau_i = (x_i, N_i, x_{N_i}, \varepsilon_i, \omega_i)$ summarize the information available to agent i - upon becoming informed, agent i observes her own type x_i, her neighbors N_i, neighbors' types x_{N_i}, her private idiosyncratic shock ε_i and her parent ω_i. Let the degree $\mathbf{d}_i$ summarize her neighborhood attribute. Each agent observes only one parent that adopts the behavior and regards the other neighbors as not exposed. The player type distribution T and degree distributions $D_{\mathbf{X}}$ are common knowledge. Importantly, agents do not observe the number of periods that have passed. Agent i is aware of N_i because there are direct links between i and each member of N_i, but agent i does not observe the whole network.

Payoff. The payoff of agent i depends on the net benefit from adoption if no other neighbor adopts, h_i, on the social effect and on the shock ε_i. For the social effect, we sum up the choices of all neighbors by weight γ.

[8]In network science, the configuration model is a method for generating random networks. In a configuration model, the network $G^{(n)}$ is drawn uniformly at random among those with a given type and degree sequence $\{(\mathbf{x}^{(n)}, \mathbf{d}^{(n)})\}_{n\in\mathbb{N}}$. It is widely used as a reference model for real-life social networks because it allows arbitrary degree sequences (Newman, 2010[46]).

[9]Later sections describe the branching process in more detail.

$$U_i(y_i, (y_j)_{j \in N_i}, \tau_i) = y_i \left(h_i + \sum_{j \in N_i} \gamma_{ij} \cdot y_j - \varepsilon_i \right), \tag{2.1}$$

where $\gamma_{ij} > 0$ measures the weight that agent i assigns to her neighbor j's choice. It summarizes the influence of player j on player i's decision making. In reality, the weight can depend on the strength of the connection between i and j, the type of the two agents, the number of neighbors agent i has, etc. So far, we assume γ_{ij} is pair-specific. The non-adoption payoff is normalized to zero by construction.

Equilibrium concept. Let $\mathcal{T}$ denote the support of agent i's information τ_i. A strategy for agent i is a mapping $s_i(\tau_i) : \mathcal{T} \to \mathcal{Y} = \{0, 1\}$ that maps i's information to the adoption decision. Let $A_i = \sum_{j \in N_i} \gamma_{ij} \cdot y_j$ denote the weighted sum of neighbors that adopt the behavior. The distribution of A_i depends on the profile of strategies $(s_i)_{i \in \mathcal{N}}$. Upon exposure, agent i forms a belief about A_i, conditioned on her information τ_i and on the fact that she adopted the behavior.[10] The belief does not change with the number of periods that have passed because agent i does not observe it.

A perfect Bayesian Nash equilibrium is a profile of strategies $(s_i)_{i \in \mathcal{N}}$ such that:

$$s_i(\tau_i) \in \arg\max_{y_i \in \mathcal{Y}} \mathbf{E}_i[U_i(y_i, \mathbf{s}_{-i}(\tau_{-i}), \tau_i) | \tau_i] = \begin{cases} 1 & \text{if } \varepsilon_i \leq h_i + \mathbf{E}_i(A_i) \\ 0 & \text{if } \varepsilon_i \geq h_i + \mathbf{E}_i(A_i) \end{cases} \quad \forall i \in \mathcal{N}, \tag{2.2}$$

where $\mathbf{s}_{-i}(\tau_{-i}) = (\mathbf{s}_j(\tau_j))_{j \in \mathcal{N} \setminus i}$. That is, each agent maximizes her expected payoff given her belief over A_i.

I maintain the following assumption about the exogenously given network.

Assumption 3. *The following hold for any n.*

a) The idiosyncratic shocks $\{\varepsilon_i\}_{i \in \mathcal{N}}$ are independently and identically distributed with strictly increasing c.d.f. F_ε and density f_ε, which is common knowledge among players.

b) For each i, $\varepsilon_i \perp (x_i, N_i)$.

[10] We focus on the payoff from adoption since we normalized the non-adoption payoff to zero.

This assumption states that the unobservable shocks ε_i are independent of each other and do not degenerate, and that ε is independent of any factors that affect link formation.

The diffusion process characterized in this section is a version of an expectation model of diffusion studied in the literature, that focus on players who care about the expected number of neighbors that will participate. (See Sadler (2020)[51] for an example.) Different from Sadler's model, we allow players to observe the neighbor types and the identity of the parent.

2.2.1 Equilibrium

This subsection focuses on the existence of equilibria in this network and demonstrates the equilibrium through a fixed-point characterization. We also propose an anonymous version of the game and show the uniqueness of the equilibrium.

Note that this is an incomplete information game due to private link and payoff information, so agents in the social network form beliefs about their neighbors' choices when they get exposed to the behavior, as mentioned in the equilibrium concept of the model.

The equilibrium approach is a standard approach to take in the network games (See Jackson and Yariv (2007)[34]. Also see Eraslan and Tang (2017)[20] for the estimation of static network models using equilibrium approach.) It requires that players have a "correct" common prior about the network formation. Canen et al. (2020)[12] take a behavioral approach to model networks games and provides an estimation method for a model of linear local interactions, which may be restrictive in application. Instead, we assume the existence of common prior and thus can demonstrate the existence of a symmetric equilibrium where a player's equilibrium beliefs only depend on anonymized information.

Recall that by Equation (2.2),

$$y_i = 1 \quad \text{iff} \quad \varepsilon_i \leq h_i + \mathbf{E}_i(A_i) = h_i + \left(\gamma_{i,\omega_i} + \sum_{j \in N_i \setminus \omega_i} \gamma_{ij} \cdot \mathbf{E}_i(y_j) \right).$$

When agent i becomes aware, she observes that ω_i has adopted, and expects neighbor $j \in$

$N_i \setminus \omega_i$ to adopt the behavior with probability $\mathbf{E}_i(y_j)$ if she adopts, where $\mathbf{E}_i(y_j)$ is the belief formed, conditioned on her information τ_i and on the fact that she adopted the behavior.

In the rest of the section, we focus on the case in which the parent node ω_i connects to any member in i's neighborhood N_i with a probability approaching zero as n approaches infinity. Thus, the belief $\mathbf{E}_i(y_j)$ is formed, conditioned on agent i's information τ_i and on the fact that she passed the information to j. We will discuss the case where ω_i connects to some members of N_i in the Appendix.

We assume the players are rational in the following sense.

Assumption 4. *(Rational Expectation) The individuals possess self-consistent expectations, that is,*

$$\mathbf{E}_i(y_j) = \mathbf{E}\left(y_j | \tau_i, \omega_j = i, i \in \mathcal{I}\right).$$

where $\mathcal{I} = \{i : \exists j \in N_i \text{ such that } y_j = 1\}$ *denotes the set of informed players.*

Assumption 4 states that the belief $\mathbf{E}_i(y_j)$ should be equal to the expected value of agent j's choice conditional on τ_i and on i being j's parent. Denote the belief $\mathbf{E}_i(y_j)$ by σ_{ij}, the expected weighted sum of adopters $\mathbf{E}_i(A_i)$ by a_i.

$$a_i(N_i, \omega_i, \sigma) = \gamma_{i,\omega_i} + \sum_{j \in N_i \setminus \omega_i} \gamma_{ij} \cdot \sigma_{ij}. \tag{2.3}$$

Agent i's action is

$$y_i = \mathbb{1}\{\varepsilon_i < h_i + a_i\}. \tag{2.4}$$

Then we show that there's a fixed-point characterization that maps σ to itself on a closed interval because σ_{li} can be written as a function of σ_{ik} for all $k \neq i$. Denote the projection by $\sigma_{li} = R_{li}\left(\sigma_{i,-i}\right)$. It is a continuous mapping from $[0,1]^{n(n-1)}$ to $[0,1]^{n(n-1)}$. We prove the existence of the equilibrium using Brouwer's fixed-point theorem.

Theorem 4. *(Existence of the equilibrium) Under Assumptions 3-4, there exists an equilib-*

rium belief σ^*. Moreover, σ^* satisfies the following self-projection $\sigma = \boldsymbol{R}(\sigma)$,

$$\sigma_{li} = R_{li}\left(\sigma_{i,-i}\right) = \sum_{N_i} F_{\varepsilon}\left(h_i + a_i(N_i, \omega_i = l, \sigma)\right) \cdot P(N_i|\omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}), \quad \forall l \neq i \quad (2.5)$$

where $a_i(N_i, \omega_i = l, \sigma) = \gamma_{il} + \sum_{j \in N_i \backslash l} \gamma_{ij}\sigma_{ij}$.

Proof. See Appendix. ☐

For estimation purposes, the rest of the paper will focus on the diffusion processes with an assumption of anonymity, that is, the equilibrium belief σ^*_{ij} depends only on the agent type x_i, x_j and neighborhood attribute $\mathbf{d}_i$, not on the identity of the nodes. Furthermore, we prove that the symmetric equilibrium is unique.

Assumption 5. *(Anonymity) The deterministic part of an agent's payoff is affected only by the characteristics (rather than actual identities) of the agent and her neighbors, specifically, $h_i = h(x_i, \mathbf{d}_i)$ and $\gamma_{ij} = \gamma(x_j, x_i, \mathbf{d}_i)$.*

The above assumption arises from the empirical literature. In social interaction models, the social effect is often characterized by the sum or the average of neighbors' decisions. An example for the first type (i.e., the sum) is $\gamma_{ij} = \gamma(x_j, x_i)$, which depends only on the types of both ends of a link. For the second type (i.e., the average), a typical example is $\gamma_{ij} = \gamma(x_j, x_i)/d_i$, where d_i is the total number of links that agent i has.

Following Assumption 5, we prove that, when two players have the same profile of individual and neighborhood characteristics, they tend to form the same beliefs for a mutual friend. For simplicity of notation, we use σ instead of σ^* to denote a Bayesian equilibrium in the following discussion.

Assumption 6. *(Anonymous beliefs) $\sigma_{ij} = \sigma_{ik}$ if $x_j = x_k$ for all $i, j, k \in \mathcal{N}$.*

Recall that $\sigma_{ij} = \mathbf{E}(y_j|\omega_j = i, \tau_i, i \in \mathcal{I})$ and $\sigma_{ik} = \mathbf{E}(y_k|\omega_k = i, \tau_i, i \in \mathcal{I})$. Given that $x_j = x_k$, on the one hand, agent j and k are observationally identical to agent i; on the other hand, by Assumption 3 and given the same parent node i, the conditional distribution of

$\mathbf{d}_j$ and $\mathbf{d}_k$ are the same. Thus, we can assume that $\sigma_{ij} = \sigma_{ik}$, because the probability that agent j or k would adopt the behavior depends on the type x_j, x_k and possible neighborhood attributes $\mathbf{d}_j$ and $\mathbf{d}_k$.

Then, we show that the equilibrium beliefs are symmetric given Assumptions 5 and 6.

Proposition 2. *(Symmetry in equilibrium beliefs) A Bayesian equilibrium σ is symmetric under Assumptions 5 and 6, that is,*

$$\sigma_{ij} = \sigma_{kl} \quad \textit{if} \quad x_i = x_k, x_j = x_l, \mathbf{d}_i = \mathbf{d}_k.$$

Proof. See Appendix. □

Proposition 2 suggests that the equilibrium beliefs depend only on the agent's and her neighbors' characteristics, not on the specific identity of players, i.e., $\sigma_{ij} = \sigma(x_j, x_i, \mathbf{d}_i)$. As a result, a_i defined in Equation (2.3) depends only on player i's individual and neighbor characteristics $(x_i, \mathbf{d}_i)$ and parent information ω_i.

One aspect of diffusion processes is that the model generally admits multiple equilibria. Some equilibria are robust to small perturbations and are therefore stable, while, for other equilibria, small perturbations lead to significant changes. With some additional restrictions, we show that a unique symmetric equilibrium belief exists.

Assumption 7. *For the shock ε and payoff parameter γ's,*

a) $F_\varepsilon(\cdot)$ is Lipschitz continuous with constant K, i.e., $f_\varepsilon(\cdot) \leq K$;

b) there exists an $\alpha \in (0,1)$ such that $K\bar{\phi} \leq \alpha$, where $\bar{\phi} = \max_{x_i,\mathbf{d}_i}\{\gamma_i \mathbf{d}_i' - \gamma_{i,\min}\}$, $\gamma_i = (\gamma(X_1, x_i, \mathbf{d}_i), ..., \gamma(X_K, x_i, \mathbf{d}_i))$ and $\gamma_{i,\min}$ is the minimum element of γ_i.

The above assumption on ε and γ limits the variation in the expected payoff. The variation cannot be too large, otherwise it affects the stability of the equilibrium and we may get multiple equilibria.

Theorem 5. *(Uniqueness of Nash equilibrium) A unique Nash equilibrium σ exists if Assumption 7 holds.*

Proof. See Appendix. □

Given that $R(\cdot)$ is upward sloping (because payoffs exhibit complementarity), the shape restrictions posted by Theorem 5 guarantee a flatter $R(\cdot)$. Together with Proposition 2, Theorem 5 implies that there exists a unique, symmetric equilibrium.

2.3 Identification

Now we discuss the identification of the model when the number of agents approaches infinity, and only a single large network is observed. In this section, we use $*$ to indicate an identifiable element. Throughout this section, we maintain Assumption 3-7 so that there is a unique, symmetric equilibrium for players' beliefs about neighbors' actions.

Suppose the researcher collects data from a single network with n individuals. For each informed agent $i \in \mathcal{I}$, the data includes the agent's type x_i, the identity of the parent ω_i, the type of the parent x_{ω_i}, the neighborhood attributes $\mathbf{d}_i$, and individual choices y_i when the diffusion process is finalized. Recall that the network is observed by the researcher since it can be recovered from historical data that demonstrates long-run relationships amongst these agents. Such a size-n network is a single, random realization of some data-generating process governed by the type and degree distributions $\left(T^{(n)}, D_{\mathbf{X}}^{(n)}\right)$. $\left(T^{(n)}, D_{\mathbf{X}}^{(n)}\right)$ converges to $(T, D_{\mathbf{X}})$ as n approaches infinity. Let $E_n(\cdot)$ denote the expectation under $\left(T^{(n)}, D_{\mathbf{X}}^{(n)}\right)$ in the following statement.

For identification purposes, this paper only discusses large cascades on networks where a substantial proportion of the population is informed during the diffusion process, i.e., the size of the informed group $|\mathcal{I}_n|$ grows linearly in n as n approaches infinity. Examples of such phenomena include the diffusion of norms and innovations, collective action, propagation of rumors, and cultural fads, where the information reaches a non-negligible percentage of the population.

Assumption 8. *(Existence of the Limit) For any* $x, x' \in \mathbf{X}$, $\mathbf{d}, \mathbf{d}' \in \mathbb{N}^K$

$$q^*(\mathbf{d}'|x', x, \mathbf{d}) \equiv \lim_{n\to\infty} \frac{1}{|\mathcal{I}_n|} \sum_j E_n \left[\mathbb{1}\left\{\mathbf{d}_i = \mathbf{d}'\right\}|x_i = x', x_{\omega_i} = x, \mathbf{d}_{\omega_i} = \mathbf{d}\right]$$

exists.

With the above assumption, we show that the asymptotic moment of equilibrium beliefs also exists. The following proposition illustrates the existence of limit $\sigma^*(x', x, \mathbf{d})$ and relates the parameters (h, γ) to asymptotic moments σ^* and q^*.

Proposition 3. *Suppose Assumption 8 holds,*

$$\sigma^*(x', x, \mathbf{d}) \equiv \lim_{n\to\infty} \frac{1}{|\mathcal{I}_n|} \sum_j E_n(y_i|x_i = x', x_{\omega_i} = x, \mathbf{d}_{\omega_i} = \mathbf{d}) \tag{2.6}$$

exists and

$$\sigma^*(x', x, \mathbf{d}) = \sum_{\mathbf{d}'} F_\varepsilon\left(h(x', \mathbf{d}') + a(x', \mathbf{d}', x_\omega = x, \sigma^*)\right) \cdot q^*(\mathbf{d}'|x', x, \mathbf{d}).$$

where $a(x', \mathbf{d}', x_\omega = x, \sigma^*) = \gamma(x, x', \mathbf{d}')(1 - \sigma^*(x, x', \mathbf{d}')) + \sum_{k=1}^K \gamma(X_k, x', \mathbf{d}')\sigma^*(X_k, x', \mathbf{d}')d'_k$.

Proof. See Appendix. □

Equation (2.6) is derived from the self-projection $\sigma = \boldsymbol{R}(\sigma)$ by replacing the equilibrium beliefs and neighborhood attributes probabilities in Equation (2.5) with identifiable asymptotic moments.

The rest of Section 2.3 illustrates the steps for identification. We maintain Assumption 3-8, so that the asymptotic moments are considered identified as probability limits. Then the payoff parameters are recovered from these asymptotic moments.

Note that, for each i that is informed, we observe player i's choice y_i, parent type x_{ω_i} and neighborhood attribute $\mathbf{d}_i$. Therefore, the conditional probability that agent i would adopt $\mathbf{E}_n(y_i|x_i, \mathbf{d}_i, x_{\omega_i}, \sigma^*)$ is identified under the asymptotics that the network size n approaches infinity. On the other hand, Equation (2.3)-(2.4) implies that $\mathbf{E}_n(y_i|\mathbf{d}_i, x_{\omega_i}, \sigma^*) =$

$F_\varepsilon\left(h_i + \gamma_{i,\omega_i} + \sum_{j\in N_i\backslash\omega_i}\gamma_{ij}\cdot\sigma^*_{ij}\right)$. By Assumption 5 and Proposition 2 that the parameters and the equilibrium beliefs depend only on the agent's and her neighborhood's characteristics, we rewrite the previous equation as

$$\mathbf{E}_n(y_i|\mathbf{d}_i, x_{\omega_i}, \sigma^*) = F_\varepsilon\left(h(x_i, \mathbf{d}_i) + \gamma(x_{\omega_i}, x_i, \mathbf{d}_i) + \sum_{j\in N(\mathbf{d}_i)\backslash\omega_i}\gamma(x_j, x_i, \mathbf{d}_i)\cdot\sigma^*(x_j, x_i, \mathbf{d}_i)\right), \tag{2.7}$$

where $N(\mathbf{d}_i)$ denotes a neighborhood that has degree $\mathbf{d}_i$, i.e., if $\mathbf{d}_i = (1, 2)$, then there are one type-1 agent and two type-2 agents in this neighborhood. The parameters to be identified are $h(\cdot,\cdot)$ and $\gamma(\cdot,\cdot,\cdot)$.

The following assumption demonstrates two possible types of restrictions that ensure identification.

Assumption 9. *There are two types of restrictions:*

a) $\gamma(x', x, \mathbf{d}) = \gamma(x', x)$ *for any* $(x', x, \mathbf{d})$*;*

b) $h(x, \mathbf{d}) = h(x)$ *and* $\gamma(x', x, \mathbf{d}) = \phi(x)w(x', x, \mathbf{d})$ *for any* $(x', x, \mathbf{d})$*, and there exist functions* $w_0(\cdot)$ *such that* $w(x', x, \mathbf{d}) = \frac{w_0(x', x)}{\sum_{j\in N(\mathbf{d})} w_0(x_j, x)}$.

Assumption 9(a) holds when the players' payoffs depend on the headcount of adopters instead of the proportion. Assumption 9(b) holds when payoffs are affected only by the weighted proportion of adopters.

The parameters are then identified with the additional restrictions.

Proposition 4. *Suppose Assumptions 3-9 hold. Then* (h, γ) *are identified from the asymptotic moments.*

Proof. See Appendix. □

Equation (2.7) implies that (h, γ) cannot be identified without further restrictions. We can see this by letting $T(x_i, \mathbf{d}_i) = \sum_{j\in N(\mathbf{d}_i)}\gamma(x_j, x_i, \mathbf{d}_i)\sigma^*(x_j, x_i, \mathbf{d}_i)$ be the weighted sum of beliefs for agents with type x_i and neighborhood attributes $\mathbf{d}_i$. Then $\mathbf{E}_n(y_i|\mathbf{d}_i, x_{\omega_i}, \sigma^*) =$

$F_\varepsilon\{h(x_i, \mathbf{d}_i) + T(x_i, \mathbf{d}_i) + \gamma(x_{\omega_i}, x_i, \mathbf{d}_i)\,(1 - \sigma^*(x_{\omega_i}, x_i, \mathbf{d}_i))\}$. Thus, even if $S(x_{\omega_i}, x_i, \mathbf{d}_i) = h(x_i, \mathbf{d}_i) + T(x_i, \mathbf{d}_i) + \gamma(x_{\omega_i}, x_i, \mathbf{d}_i)\,(1 - \sigma^*(x_{\omega_i}, x_i, \mathbf{d}_i))$ can be identified as a result of Assumption 3(a) such that $F_\varepsilon(\cdot)$ is strictly monotonic, we still cannot identify h and γ for any function $h + T$. This is because there would always exist a weight function γ such that the equality holds. This is achieved by redistributing weight γ and adjusting the add-on h, conditioning on each $(x, \mathbf{d})$.

To see it more clearly, I reshape the original expression for function $S(\cdot)$ (see Appendix for the proof),

$$S^*(x_j, x_i, \mathbf{d}_i)(1+Y^*(x_i, \mathbf{d}_i)) - X^*(x_i, \mathbf{d}_i) = h(x_i, \mathbf{d}_i) + \gamma(x_j, x_i, \mathbf{d}_i)(1-\sigma^*(x_j, x_i, \mathbf{d}_i))(1+Y^*(x_i, \mathbf{d}_i)),$$

where $X^*(x_i, \mathbf{d}_i) = \sum_{j \in N(\mathbf{d}_i)} S^*(x_j, x_i, \mathbf{d}_i) \frac{\sigma^*(x_j, x_i, \mathbf{d}_i)}{1-\sigma^*(x_j, x_i, \mathbf{d}_i)}$, $Y^*(x_i, \mathbf{d}_i) = \sum_{j \in N(\mathbf{d}_i)} \frac{\sigma^*(x_j, x_i, \mathbf{d}_i)}{1-\sigma^*(x_j, x_i, \mathbf{d}_i)}$. The above equation illustrates that the identification is not achievable even if $S^*(x_j, x_i, \mathbf{d}_i)$ and $\sigma^*(x_j, x_i, \mathbf{d}_i)$ are identified for all combinations of $(x_j, x_i, \mathbf{d}_i)$, because, for each $(x_i, \mathbf{d}_i)$, we have K equalities but $K + 1$ unknowns to be identified. Therefore, we propose further restrictions that guarantee the identification of parameters (h, γ) under the condition that σ^* is identified by Proposition 3.

2.4 Two-Step M-Estimator

We propose a two-step m-estimator for parameters in individual payoffs. In the first step, we estimate σ using asymptotic moments. Then, by plugging in the estimated σ, we estimate the coefficient θ using an M.L.E. estimator.

2.4.1 First Step: the Equilibrium Belief σ

We begin this section by showing that the asymptotic moments σ in Equation (2.6) can be consistently estimated using sample averages across individuals from a single network as the number of individuals $n \to \infty$. Moreover, we prove the root-n consistency of this estimator.

Proposition 5. *Define estimator of the nuisance parameter* σ *by*

$$\hat{\sigma}(x', x, \mathbf{d}) = \frac{\sum_{i \in \mathcal{I}_n} y_i \mathbb{1}\{x_i = x'\} \iota_{\omega_i}(x, \mathbf{d})}{\sum_{i \in \mathcal{I}_n} \mathbb{1}\{x_i = x'\} \iota_{\omega_i}(x, \mathbf{d})}.$$

If Assumption 3-6 holds, then we have $||\hat{\sigma} - \sigma^*|| = O_p\left(\frac{1}{\sqrt{n}}\right)$, *i.e.,*

$$\sup_{\substack{x, x' \in \mathbf{X} \\ m \in \mathbf{M}}} |\hat{\sigma}(x', x, \mathbf{d}) - \sigma^*(x', x, \mathbf{d})| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

where σ^* *is the true value of the equilibrium belief.*

Proof. See Appendix. □

Proposition 5 states that the average of the sample analog is a uniformly root-n consistent estimator for the equilibrium belief σ. The proof relies on the fact that players' decisions are conditionally independent of each other given that researchers observe the types and degree $(x', x, \mathbf{d})$. Combined with the definition of *large cascades*, that the size of the informed group $|\mathcal{I}_n|/n \to \alpha_0$ for some $\alpha_0 > 0$, we demonstrate the root-n consistency of the first-step estimator $\hat{\sigma}$.

2.4.2 Second Step: the Maximum Likelihood Estimator

We can sort the nodes by period. The diffusion process starts with a single individual at time 0. Each individual makes an irreversible decision in a single period; if she decides to adopt, she will pass the information to her neighbors. Let N_t be the set of individuals that make decisions at period t. The process is $\{N_0, N_1, ..., N_T\}$. The T is the last period of the observation. It does not necessarily have to be the termination of the diffusion process.

Suppose we observe a realization of the process $\{y_{ti}, x_{ti}, \mathbf{d}_{ti}, \omega_{ti}\}$ for $i = 1, ..., n_t \in N_t$ and $t = 1, ..., T$. Denote the observed information about the players who adopt in the first t periods by $\mathcal{H}_t$. Then $\mathcal{H}_t = \{y_{t'i}, x_{t'i}, \mathbf{d}_{t'i}, \omega_{t'i}\}_{\substack{i \in N_{t'} \\ t'=1,...,t}}$. We can uncover the payoff function by finding a parameter θ that maximizes the likelihood function L defined by

$$
\begin{aligned}
& L\left(\mathcal{H}_T \middle| y_0 = 1, x_0, \mathbf{d}_0, \theta\right) \\
&= \left(\prod_{i \in N_T} P(y_{Ti} | x_{Ti}, \mathbf{d}_{Ti}, \omega_{Ti}, \theta) \right) \cdot \pi\left(\{x_{Ti}, \mathbf{d}_{Ti}, \omega_{Ti}\}_{i \in N_T} \middle| \mathcal{H}_{T-1} \right) \cdot L\left(\mathcal{H}_{T-1} \middle| y_0 = 1, x_0, \mathbf{d}_0, \theta\right) \\
&= \prod_{t=1}^{T} \left\{ \left(\prod_{i \in N_t} P(y_{ti} | x_{ti}, \mathbf{d}_{ti}, \omega_{ti}, \theta) \right) \cdot \pi\left(\{x_{ti}, \mathbf{d}_{ti}, \omega_{ti}\}_{i \in N_t} \middle| \mathcal{H}_t \right) \right\}
\end{aligned}
\tag{2.8}
$$

where θ is the parameter of interest (h, γ) over all types and degrees. The first equality demonstrates the likelihood function in an iterative form, and the second equality is an expansion. Thus, the likelihood function consists of two parts: the conditional probability of period t players' decisions and the forward function π.

We can easily calculate the conditional probability of player i's decision, as in the previous sections $P(y_i = 1 | x_i, \mathbf{d}_i, \omega_i, \theta, \sigma) = F_\varepsilon \left\{ h_i + \gamma_{i,\omega_i}(1 - \sigma_{i,\omega_i}) + \sum_{j \in N(\mathbf{d}_i)} \gamma_{ij}\sigma_{ij} \right\}$. For types with finite and discrete support $\mathbf{X} = \{X_1, ..., X_K\}$, after categorizing the terms in the summation by types,

$$P(y_i = 1 | x_i, \mathbf{d}_i, \omega_i, \theta, \sigma) = F_\varepsilon \left(Z_i(x_i, \mathbf{d}_i, \omega_i, \sigma)' \theta_i \right),$$

where $\theta_i = \theta(x_i, \mathbf{d}_i) = (h(x_i, \mathbf{d}_i), \gamma(X_1, x_i, \mathbf{d}_i), ..., \gamma(X_K, x_i, \mathbf{d}_i))'$, and $Z_i(x_i, \mathbf{d}_i, \omega_i, \sigma) = (1, \sigma(X_1, x_i, \mathbf{d}_i)d_{i1} + \mathbb{1}\{X_1 = x_{\omega_i}\}(1 - \sigma(X_1, x_i, \mathbf{d}_i)), ..., \sigma(X_K, x_i, \mathbf{d}_i)d_{iK} + \mathbb{1}\{X_K = x_{\omega_i}\}(1 - \sigma(X_K, x_i, \mathbf{d}_i)))'$.

Now we show that the forward function π in Equation (2.8) does not depend on θ in the limit $(n \to \infty)$, that is, $\pi\left(\{x_{ti}, \mathbf{d}_{ti}, \omega_{ti}\}_{i \in N_t} \middle| \mathcal{H}_t \right)$ is not a function of θ.

First of all, since $\{x_{ti}, \omega_{ti}\}_{i \in N_t}$ are jointly determined by information contained in the history $\{y_{t-1,i}, x_{t-1,i}, \mathbf{d}_{t-1,i}\}_{i \in N_{t-1}}$, the forward function π degenerates to $\pi\left(\{\mathbf{d}_{ti}\}_{i \in N_T} \middle| \mathcal{H}_t \right)$.

The distribution of $\mathbf{d}_{ti}$ is restricted by the following conditions:

1. one link of node i connects to a type-x_{ω_i} player since $\omega_i \in N_i$,

2. $\mathbf{d}_{\omega_i}$ affects the distribution of $\mathbf{d}_i$ because i may be connected to some $j \in N_{\omega_i}$,

3. $\omega_i \in \mathcal{I}$ thus $\omega_{\omega_i} \in \mathcal{I}$.

The first two do not depend on σ since they are determined by the type and degree distributions that govern the network generating process. Only the third condition is potentially related to the diffusion process.

When agent l does not observe the identity of her parent, her beliefs about her neighbors' degrees are affected by the viral belief distortion. However, when the identity of ω_l is observed, viral belief distortion affects only l's belief about ω_l's degree, not her beliefs about other neighbors' degrees. In sum, $P(\mathbf{d}_i|\omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I})$ does not change with θ.

The core idea of the proof is to approximate the local network structure with a branching process. The branching process is defined by the diffusion process from a "seed" player in the network. The "seed" player's neighbors are the first generation, and the neighbors' neighbors are the second, and so on. For each informed agent i, there is a unique period t_i in which agent i decides to adopt. This t_i corresponds to the generation in which agent i lies in in the branching process.

In any finite graph, the offspring distributions in this diffusion process are not independent, and they change over time. However, Lemma 1 of Sadler (2020)[51] shows that these complications are asymptotically insignificant: as n grows, the local structure of the configuration model converges to that of a branching process. This result relies on the configuration model of random graphs, which takes a uniform random draw among all graphs G with a given type and degree sequence for the nodes in the network.

With the isomorphic structure in asymptotics, the neighborhood attributes of players informed in N_t depend only on the types and degrees of adopters in period $t-1$ and are independent of the history $\mathcal{H}_{t-2}$. For example, for period T, $\pi\left(\{x_{Ti}, \mathbf{d}_{Ti}, \omega_{Ti}\}_{i \in N_T} \middle| \mathcal{H}_{T-1}\right) = \pi\left(\{\mathbf{d}_{Ti}\}_{i \in N_T} \middle| \{y_{ti}, x_{ti}, \mathbf{d}_{ti}\}_{i \in N_{T-1}}\right)$.

Therefore, θ affects the likelihood function $L(\cdot)$ only through the conditional probability of adoption by each informed player, $L\left(\mathcal{H}_T \middle| y_0 = 1, x_0, \mathbf{d}_0, \theta\right) = \prod_{i \in \mathcal{I}} P(y_{ti}|x_{ti}, \mathbf{d}_{ti}, \omega_{ti}, \theta)$.

The second-step MLE objective function is

$$\hat{Q}_n(\theta, \hat{\sigma}) = \frac{1}{|\mathcal{I}_n|} \sum_{i \in \mathcal{I}_n} \{y_i \log F_\varepsilon (Z_i(\hat{\sigma})'\theta_i) + (1 - y_i) \log (1 - F_\varepsilon (Z_i(\hat{\sigma})'\theta_i))\} \tag{2.9}$$

where $Z_i(\hat{\sigma}) = Z_i(x_i, \mathbf{d}_i, \omega_i, \hat{\sigma})$, and $\mathcal{I}_n$ is the group of agents who are connected with each other in a giant component of this network index by n. Partition the set of individuals $\mathcal{I}_n$ into $\mathcal{I}_{n,k} = \{i \in \mathcal{I}_n : x_i = X_k\}$. This is equivalent to finding the $\hat{\theta}$ that maximizes $\hat{Q}_{n,k}(\theta, \hat{\sigma}) = \frac{1}{|\mathcal{I}_{n,k}|} \sum_{i \in \mathcal{I}_{n,k}} \{y_i \log F_\varepsilon (Z_i(\hat{\sigma})'\theta) + (1 - y_i) \log (1 - F_\varepsilon (Z_i(\hat{\sigma})'\theta))\}$, where $\mathcal{I}_{n,k}$ is the group of agents within this giant component that has the same type X_k.

Assumption 10. *The following conditions are needed for the consistency of the estimator for θ:*

a) $\theta \in \Theta$, Θ is compact;

b) $\liminf_n \frac{1}{|\mathcal{I}_{n,k}|} Z'Z > 0$, where $Z = (Z_i(\sigma_0)')_{i \in \mathcal{I}_{n,k}}$ for all k.

Compactness of Θ is Assumption 10(a) guarantees the maximum of the probability limit of the MLE objective function. Assumption 10(b) requires that the $(K+1) \times (K+1)$ matrix $Z'Z$ does not degenerate when n approaches infinity and thus the parameters in the model are identified.[11]

The following theorem states the consistency of the second-step estimator $\hat{\theta}$.

Theorem 6. *(Consistency) Suppose Assumptions 3-10 hold. Then $\hat{\theta} \xrightarrow{p} \theta_0$.*

Proof. See Appendix. □

By Proposition 5, for any θ in the compact set Θ (e.g., a closed interval $[a, b]^{K+1}$), $\hat{\theta}$ is consistent if the matrix $Z'Z$ does not degenerate when n approaches infinity.

In sum, this two-step approach that applies to the diffusion model differs significantly from a static network model. (See Eraslan and Tang (2017)[20] for a two-step approach that

[11] We provided sufficient conditions for Assumption 10(b) in Section 2.3.

applies to static network games.) We have to impose restrictions to the network formation due to possible selectivity in the diffusion games, as not everyone in the network is exposed to the new information.

2.5 Asymptotic Results

In this section, we briefly derive the asymptotic property of the two-step estimator. We maintain Assumptions 3-10 throughout this section.

Notice that we already proved the following conditions in Proposition 5 and Theorem 6, respectively,

a) $||\hat{\sigma} - \sigma_0|| = O_p(1/\sqrt{n})$;

b) $\hat{\theta} - \theta_0 \xrightarrow{p} 0$.

Let $f_i(\theta, \sigma) = y_i \log F_\varepsilon(Z_i(\sigma)'\theta) + (1-y_i)\log(1-F_\varepsilon(Z_i(\sigma)'\theta))$, and $W_i = E\left[\nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0)|Y\right]$, the estimator is asymptotically normal, as stated in the following theorem.

Theorem 7. *(Asymptotic Normality) Suppose Assumptions 3-10 hold.*

$$\sqrt{|\mathcal{I}_{n,k}|}\ \Sigma_{n,k}^{-1}(\hat{\theta}_k - \theta_k^0) \xrightarrow{d} \mathcal{N}(0, I_{K+1}), \quad k = 1, ..., K \tag{2.10}$$

where

$$\Sigma_{n,k} = \Gamma_{n,k}^{-1}\Omega_{n,k}\Gamma_{n,k}^{-1},$$

$$\Gamma_{n,k} = \frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} E\left[\nabla_{\theta\theta} f_i(\theta_0, \sigma)\middle| Y\right],$$

$$\Omega_{n,k} = \frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} Var\left(\nabla_\theta f_i(\theta_0, \sigma_0)\right) + \frac{1}{|\mathcal{I}_{n,k}|}\left(\sum_{i\in\mathcal{I}_{n,k}} W_i\right) Var(\hat{\sigma}) \left(\sum_{i\in\mathcal{I}_{n,k}} W_i\right)^\top.$$

Proof. See Appendix. ∎

2.6 Extension to Players Who can Defer Their Decisions

In this section, we extend our discussion to players who can defer their adoption decisions. We assume, in the previous section, that agents make decisions when they learn about the new product or new program. However, this might be restrictive if people can postpone their adoption decisions, which happens especially when agents observe their neighbors actions and are allowed to have more time for making decisions.

We assume agents who have already learned about the new product but not yet adopted can adopt the product in a later period. To be more specific, we assume the decision to participate is irreversible, only those who did not participate are allowed to switch.

This new setup differs from the previous setup because there are stronger interdependence among agents' decisions. In the previous setup, correlated individual choices derives from strategic interaction among agents due to their types and positions in the network, nevertheless they still make conditional independent decisions because they do not observe other neighbors' decisions, except their parents. As agents are allowed to observe all their neighbors' decisions, the complementarity can lead to correlated individual choices, and the correlation can bring difficulty to the identification of social effects.

This section describes the unique equilibrium under the new setup, and provide additional assumptions to identify the payoff parameters. Same as in the previous setup, the variation in the neighborhood of different nodes makes it possible for us to have identification with one single formation of a network instead of several networks with the same network formation process.

2.6.1 Model and Equilibrium

We maintain the same assumptions for *players*, *network* and *payoff* in Section 2.2. In addition, we impose the following new model assumptions.

Diffusion process. The game characterizes the diffusion of one particular behavior. Initially none of the agents has been exposed to the behavior and only a few players are aware of it. In each following period, those who are aware decide whether to adopt the behavior, and once a player adopts, her neighbors become aware. The decision to adopt is irreversible, that is, players who already adopt the behavior could not change their decision. Individual i's choice in period t is denoted by $y_{it} \in \{0,1\}$. $y_{it} = 1$ if agent i adopts the behavior, $y_{it} = 0$ if agent i denies. The space of all binary choices in period t is $\boldsymbol{y_t} = (y_{1t}, ..., y_{nt})^\top$. Denote the final status of player i, *i.e.*, player i's choice when the diffusion process stops, by y_i.

Information. Let $\tau_i = (x_i, N_i, x_{N_i}, \varepsilon_i, y_{N_i,t})$ summarize the information available to agent i - upon becoming informed, agent i observes her own type x_i, her neighbors N_i, neighbors' types x_{N_i}, her private idiosyncratic shock ε_i and her neighbors' decisions $y_{N_i,t} = \{y_{jt} \in \{0,1\} : j \in N_i\}$. Let the degree $\mathbf{d}_i$ summarize her neighborhood attribute. The player type distribution T and degree distributions $D_{\mathbf{X}}$ are common knowledge. Importantly, agents do not observe the number of periods that have passed. Agent i is aware of N_i because there are direct links between i and each member of N_i, but agent i does not observe the whole network.

We maintain the same *equilibrium concept* as that in Section 2.2. Assumption 3 remains for the new setup.

Agents in the social network form beliefs about their neighbors' choices because of incomplete information due to private link and payoff information.

By Equation (2.2), $y_{it} = 1$ if and only if $\varepsilon_i \leq h_i + \mathbf{E}_i(A_{it})$, where

$$\mathbf{E}_i(A_{it}) = \sum_{\substack{j \in N_i \\ y_{jt}=1}} \gamma_{ij} + \sum_{\substack{k \in N_i \\ y_{kt} \neq 1}} \gamma_{ik} \cdot \mathbf{E}_i(y_k).$$

When agent i becomes aware, she expects those neighbor who have not adopted to adopt the behavior with probability $\mathbf{E}_i(y_j)$ if she adopts, where $\mathbf{E}_i(y_j)$ is the belief formed, conditioned on her information τ_i and on the fact that she adopted the behavior.

We assume the players are rational in the following sense.

Assumption 11. *(Rational Expectation) The individuals possess self-consistent expecta-*

tions, that is,

$$\mathbf{E}_i(y_j) = \mathbf{E}\left(y_j | \tau_i, y_i = 1, g_{li} = 1, i \in \mathcal{I}\right).$$

where $\mathcal{I} = \{i : \exists j \in N_i$ *such that* $y_j = 1\}$ *denotes the set of informed players.*

Denote the belief $\mathbf{E}_i(y_j)$ by σ_{ij}, the expected weighted sum of adopters $\mathbf{E}_i(A_{it})$ by a_{it}, then $a_{it}(\sigma) = \sum_{\substack{j \in N_i \\ y_{jt}=1}} \gamma_{ij} + \sum_{\substack{k \in N_i \\ y_{kt} \neq 1}} \gamma_{ik} \cdot \sigma_{ik}$.

Recall Equation (2.2), in a non-cooperation game, player i's decision y_{it}, given her expectation for other players' choice, is equal to 1 if $\varepsilon_i < h_i + a_{it}(\sigma)$ in period t. Therefore, a player adopts the behavior if the right hand side of the inequality is large enough. Since each element in vector $\boldsymbol{y_t}$ increases in t, the expected number of neighbors who adopt $a_{it}(\sigma)$ is also increasing in t. I denote the maximum of $a_{it}(\cdot)$ by a_i such that

$$a_i(\sigma) = \lim_{t \to \infty} a_{it}(\sigma) = \sum_{\substack{j \in N_i \\ y_j = 1}} \gamma_{ij} + \sum_{\substack{k \in N_i \\ y_k \neq 1}} \gamma_{ik} \cdot \sigma_{ik}. \tag{2.11}$$

Thus, players' choice function can also be denoted by its matrix form.

Matrix form. Let $\mathcal{E} = (\varepsilon_1, ..., \varepsilon_n)'$, $\mathcal{H} = (h_1, ..., h_n)'$, $\mathcal{Y} = (y_1, ..., y_n)'$, $\Gamma = [\gamma_{ij}]_{n \times n}$, and $\Sigma = [\sigma_{ij}]_{n \times n}$, then

$$\mathcal{Y} = \mathbb{1}\{\mathcal{E} < \mathcal{H} + (\Gamma \circ G)\mathcal{Y} + (\Gamma \circ G \circ \Sigma)(1 - \mathcal{Y})\}, \tag{2.12}$$

where $A \circ B$ is the Hadamard product of matrices A and B. Note that $\mathcal{H}$ is an unknown vector, $\mathcal{Y}$ is observable, Σ is unknown, $\Gamma \circ G$ is the weighted adjacency matrix. $\mathcal{H}$ and Γ are our coefficients of interest.

Then we show that there's a fixed-point characterization that maps σ to itself on a closed interval because σ_{li} can be written as a function of σ_{ik} for all $k \neq i$. Denote the projection by $\sigma_{li} = R_{li}(\sigma_{i,-i})$. It is a continuous mapping from $[0,1]^{n(n-1)}$ to $[0,1]^{n(n-1)}$. We prove the existence of the equilibrium using Brouwer's fixed-point theorem.

Theorem 8. *(Existence of the equilibrium) Under Assumptions 3 and 11, there exists an* *equilibrium belief* σ^**. Moreover,* σ^* *satisfies the following self-projection* $\sigma = \boldsymbol{R}(\sigma)$*,*

$$\sigma_{li} = R_{li}\left(\sigma_{i,-i}\right) = \sum_{N_i} E_Y\left[F_\varepsilon\left(h_i + a_i(\sigma)\right)\right] \cdot P(N_i|y_l = 1, g_{li} = 1, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}) \tag{2.13}$$

$\forall l \neq i$*, where* $a_i(\sigma) = \gamma_{il} + \sum_{j\in N_i \setminus l} \gamma_{ij} \cdot \max\{y_j, \sigma_{ij}\}$*.*

Proof. See Appendix. □

We maintain Assumptions 5 and 6 from the previous setup. Then we show that the equilibrium beliefs when agents can defer their decisions are still anonymous and symmetric.

Proposition 6. *(Symmetry in equilibrium beliefs) A Bayesian equilibrium* σ *is symmetric under Assumptions 3, 5, 6, and 11, that is,*

$$\sigma_{ij} = \sigma_{kl} \quad \textit{if} \quad x_i = x_k, x_j = x_l, \mathbf{d}_i = \mathbf{d}_k.$$

Proof. The proof is similar to the proof of Proposition 2. □

Thus, the equilibrium beliefs still depend only on the agent's and her neighbors' characteristics, not on the specific identity of players, i.e., $\sigma_{ij} = \sigma(x_j, x_i, \mathbf{d}_i)$. We further assume Assumption 7 still holds to ensure the uniqueness of the symmetric equilibrium belief.

Theorem 9. *(Uniqueness of Nash equilibrium) A unique Nash equilibrium* σ *exists if Assumption 3, 5, 6,11, and 7 holds.*

Proof. See Appendix. □

2.6.2 Estimation

In this section I discuss the estimation of parameters (h, γ) from a sequence of networks, where the number of agents in the network goes to infinity. Suppose econometricians observe *i*) outcomes y_i when the diffusion terminates, *ii*) agent characteristics x_i, and *iii*) the network G. It takes two steps to investigate the identification of the payoff function. In the first step, I show that there exists a unique solution σ^* which depends only on the types of agents. Then I argue that (h, γ) are identified under some further assumptions.

I propose two ways to estimate $\{\sigma_{ij}\}$.

- *Method 1*: For any given (h, J), solve $\{\sigma_{ij}\}$ as a function of (h, J). Estimate (h^*, J^*) that maximizes the likelihood function that I'm going to discuss in section 4.3. Then find the corresponding $\{\hat{\sigma}_{ij}\}$ with the estimated $(\hat{h}, \hat{J})$.

  This method relays heavily on computation. In addition, as we will use M.L.E to estimate parameters in the discrete choice model, expectations as a function of parameters are likely to incur more computational difficulties when deriving the asymptotics. Therefore a second method is proposed.

- *Method 2*: Recall that σ_{ij} is a type-X_i player's belief of the probability of her type-X_j neighbor adopting if that neighbor is exposed to the behavior.

  For now, we assume consistency of the following estimator:

  Assumption 12. *For any characteristics $X_i, X_j \in \Theta$,*

  $$\hat{\sigma}(x', x, \mathbf{d}) = \frac{\sum_{i \neq j} y_i \mathbb{1}\{x_i = x'\} \iota_i(x, \mathbf{d}) g_{ij}}{\sum_{i \neq j} \mathbb{1}\{x_i = x'\} \iota_i(x, \mathbf{d}) g_{ij}} \xrightarrow{p} \sigma(x', x, \mathbf{d}). \tag{2.14}$$

  Therefore $\{\sigma_{ji}\}$ can be estimated by $\{\hat{\sigma}_{ji}\}$. The consistency is proved in the next subsection.

We impose the following assumption.

Assumption 13. *(Existence of the Limit) For any $x, x' \in \mathbf{X}$, $\mathbf{d}, \mathbf{d}' \in \mathbb{N}^K$*

$$q^*(\mathbf{d}'|x', x, \mathbf{d}) \equiv \lim_{n \to \infty} \frac{1}{|\mathcal{I}_n^2|} \sum_{j \neq i} E_n \left[\mathbb{1}\left\{ \mathbf{d}_j = \mathbf{d}' \right\} | x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1 \right]$$

exists, where $|\mathcal{I}_n^2| = |\mathcal{I}_n|(|\mathcal{I}_n| - 1)$.

With the above assumption, we show that the asymptotic moment of equilibrium beliefs also exists. The following proposition illustrates the existence of limit $\sigma^*(x', x, \mathbf{d})$ and relates the parameters (h, γ) to asymptotic moments σ^* and q^*.

Proposition 7. *Suppose Assumption 13 holds,*

$$\sigma^*(x', x, \mathbf{d}) \equiv \lim_{n\to\infty} \frac{1}{|\mathcal{I}_n^2|} \sum_{j\neq i} E_n(y_j | x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1) \tag{2.15}$$

exists and

$$\sigma^*(x', x, \mathbf{d}) = \sum_{\mathbf{d}'} \left(\sum_{y_{N_i}} F_\varepsilon \left(h(x', \mathbf{d}') + a(x', \mathbf{d}', x, \sigma^*) \right) \right) \cdot q^*(\mathbf{d}' | x', x, \mathbf{d}).$$

where $a(x', \mathbf{d}', x, \sigma^*) = \gamma(x, x', \mathbf{d}') + \sum_{j \in N_i \setminus l} \gamma(x_j, x', \mathbf{d}')(y_j + (1 - y_j)\sigma^*(x_j, x', \mathbf{d}')$.

Proof. See Appendix. □

Equation (2.15) is derived from the self-projection $\sigma = \boldsymbol{R}(\sigma)$ by replacing the equilibrium beliefs and neighborhood attributes probabilities in Equation (2.13) with identifiable asymptotic moments. Then under the following mentioned assumptions, the asymptotic moments are considered identified as probability limits, and the payoff parameters are recovered from these asymptotic moments.

Proposition 8. *The unknown components* (h, γ, σ) *are identified under Assumption 3, 5, 6, 7, 11, and 13.*

Proof. See Appendix. □

The issue with the new two-step m-estimator lies in the first step when we estimate σ using asymptotic moments. We need additional assumptions to show that the asymptotic moments σ in Equation (2.15) can be consistently estimated using sample averages across individuals as the number of individuals $n \to \infty$, because deferred decisions allow agents to observe their neighbors and perform more correlated actions. Therefore, we assume that certain asymptotic properties holds.

Assumption 14 (Asymptotic Uncorrelation)**.** *For any $x, x' \in X$ and $\mathbf{d}, \mathbf{d}' \in M$,*

(i) $C_n\left(\iota_i(x,\mathbf{d}), \iota_i(x,\mathbf{d}')\right) \to 0$ *as* $n \to \infty$*;*

(ii) $C_n\left(\iota_i(x,\mathbf{d})\iota_j(x,\mathbf{d}')g_{ij}, \iota_k(x,\mathbf{d})\iota_l(x,\mathbf{d}')g_{kl}\right) \to 0$ *as* $n \to \infty$ *if* $\{i,j\} \cap \{k,l\} = \phi$*;*

(iii) $V_n[y_i\iota_i(x,m)]$ *and* $V_n[\iota_i(x,\mathbf{d})\iota_j(x,\mathbf{d}')g_{ij}]$ *exist for all* n *and are both* $o(1/n)$*.*

Proposition 9. *Define estimator of the nuisance parameter* σ *by*

$$\hat{\sigma}(x', x, \mathbf{d}) = \frac{\sum_{i \neq j} y_i \mathbb{1}\{x_i = x'\}\iota_i(x,\mathbf{d})g_{ij}}{\sum_{i \neq j} \mathbb{1}\{x_i = x'\}\iota_i(x,\mathbf{d})g_{ij}}$$

If Assumption 3-6 holds, then we have $||\hat{\sigma} - \sigma^*|| = O_p\left(\frac{1}{\sqrt{n}}\right)$*, i.e.,*

$$\sup_{\substack{x,x' \in \mathbf{X} \\ m \in \mathbf{M}}} |\hat{\sigma}(x', x, \mathbf{d}) - \sigma^*(x', x, \mathbf{d})| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

where σ^* *is the true value of the equilibrium belief.*

Proof. See Appendix. $\square$

Proposition 9 states that the average of the sample analog is a uniformly root-n consistent estimator for the equilibrium belief σ if the network exhibits asymptotic uncorrelation. This is a relatively weak condition if the network is generated from a configuration model, as we assume throughout this paper.

We apply maximum likelihood estimation in the second step. For those players who are informed, define $a_i(y_{N_i}) = \sum_{j \in N_i \backslash l} \gamma_{ij} \max\{y_j, \sigma_{ij}\}$ Then the log likelihood function is

$$L = \sum_{i \in \mathcal{I}_n} y_i \log\left[F(h_i + a_i(y_{N_i}))\right] + (1 - y_i)\log[1 - F(h_i + a_i(y_{N_i}))] \tag{2.16}$$

We maintain Assumption 10. The following theorem states the consistency of the second-step estimator $\hat{\theta}$.

Theorem 10. *(Consistency) Suppose Assumptions 3, 5, 6, 7, 11, 13, and 14 hold. Then* $\hat{\theta} \xrightarrow{p} \theta_0$.

Proof. The proof is similar to the proof of Theorem 6. $\square$

2.7 Simulation

2.7.1 Data Generating Process

In this section, we present the simulation results of our method. Suppose there are two types of players, type 0 and type 1. The payoff of each agent in the simulation is parameterized as

$$U_i(y_i, y_{j:j\in N_i}, \tau_i) = y_i \left(h_i + \sum_{j\in N_i} \gamma_{ij} y_j - \varepsilon_i \right),$$

where $h_0 = 0.35$, $\gamma_{00} = 0.2$, $\gamma_{01} = 0.15$, and $h_1 = 0.65$, $\gamma_{10} = 0.05$, $\gamma_{11} = 0.10$ (for simplicity of notation, the subscripts denote the types of agents, not their identities). The idiosyncratic shock ε_i follows some known distribution. The maximum number of links a player can have is three. Note that, by the definition of $\bar{\phi}$ in Assumption 7, $\bar{\phi} = 0.55$. We are interested in the personal characteristics h and the peer effects γ for both types.

There are two shape restrictions according to Theorem 5:

a) $f_\varepsilon(\varepsilon) \le K$ for any ε;

b) $K\bar{\phi} \le \alpha$ for some $\alpha < 1$.

For example, if ε follows a uniform distribution $U(0, 1)$, then $K = 1$, $\alpha = 0.55$. In addition, if ε follows a normal distribution, then $K = 1/\sqrt{2\pi}\sigma = 0.3989/\sigma$, where σ is the standard deviation. Thus $\sigma = 0.3$ would satisfy the second restriction above.

In the simulated networks, individual types x_i are drawn independently from the support $\mathbb{X}$, with probability 0.4 that it is a type-0 agent and 0.6 that it is a type-1 agent. Undirected links are formed, with the maximum degree of all nodes equal to 3. The generation of the network follows a configuration model as follows. We assign a degree sequence to each vertex following the degree distribution (the probability of having 1, 2, and 3 neighbors is (0.1, 0.4, 0.5) for a type-0 agent, and (0.2,0.3,0.5) for a type-1 agent). Then we draw a degree sequence from the following distribution: with probability 0.50, a type-0 stub connects to a type-1 stub, and, with probability 0.35, a type-1 stub connects to a type-0 stub. The sum of 1-0 stubs is equal to the sum of 0-1 stubs, which is a necessary condition for the existence of the network graph with the assigned type and degree sequence. The configuration model generated the network by taking a uniform draw from all possible simple networks with these pre-defined type and degree sequences.

Only large cascades that provide enough observations are studied. We consider networks with 200,500, and 800 nodes, in which the largest connected component should consist of more than half of the nodes. In the appendix, we attached simulation results from networks with 1000 or 2000 nodes. The seed players are drawn from the largest component of this network. The resulting diffusion process should expose more than a quarter of the agents to the new product. For each sample size n considered, we simulate $r = 1$, 5, and 10 independent sample networks of size n in order to have more observations[12]. We repeat each simulation $S = 50$ times, the data observed from each player i who is informed consists of characteristics x_i, neighborhood attributes $\mathbf{d}_i$, the identity of i's parent, and the choice y_i. Table 2.1 illustrate the diffusion results from a network where a large cascade exists.

The individual choices under the symmetric pure-strategy Bayesian equilibrium are simulated using the following steps. First, we generate 500 networks of size n and simulate diffusion processes on generated networks using an initial belief (e.g., $\sigma = 0.6$ for all types). Calculate the realized probability that a neighbor adopts from the simulation result. Next, plug this probability into individuals' beliefs and simulate diffusion processes again. Repeat this operation until the probability converges. The convergence is guaranteed and the final simulated probability is the individual's true belief about her neighbors' choices, by Theorem 5 (see the proof of Theorem 5 in Appendix 2.10 for details). Finally, we simulate diffusion processes using the true belief on a new batch of networks. We estimate the payoff

[12]We take this approach only because it is difficult to generate a very large network and to find the true equilibrium belief of that network. Note that, in practice, we can estimate the equilibrium belief from observations. This is an issue only when performing simulations. In later tables, we can also find that the bias of the estimator is similar for five networks each with 2000 nodes and ten networks each with 1000 nodes.

parameters from the individual choices using the two-step m-estimator.

I study two scenarios for the idiosyncratic shock ε_i: (a) the shock follows a uniform distribution $U(0,1)$; and (b) the shock follows a normal distribution $N(0.5, 0.3)$. The distributions are chosen so that $\max(f_\varepsilon(\cdot)) \cdot \bar{\phi} < 1$, as stated in Theorem 5.

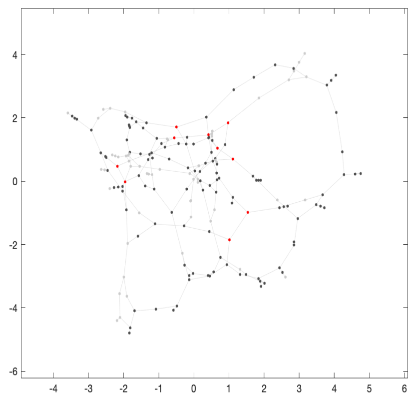
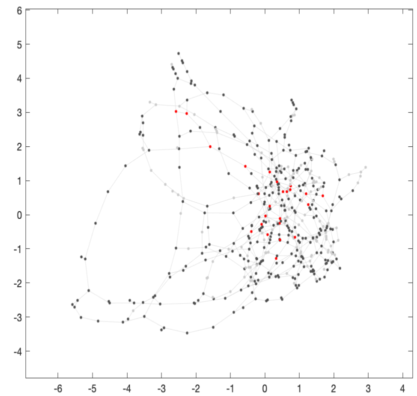
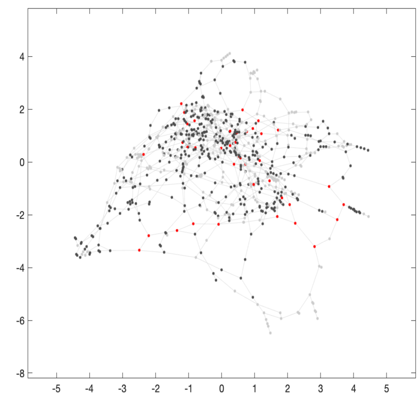
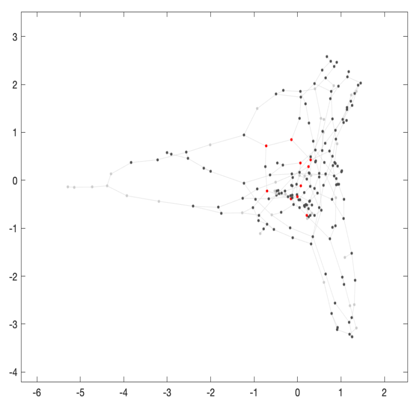
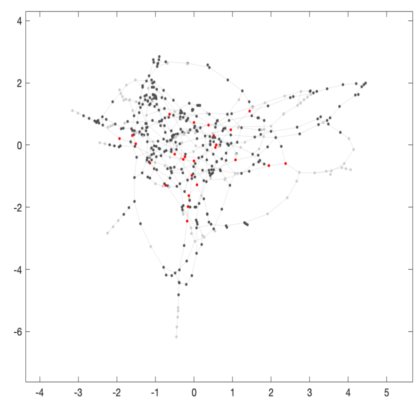
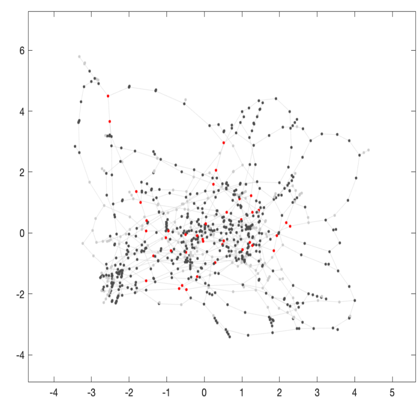
	N = 200	500	800
Uniform			
Normal			

Table 2.1: Diffusion Results

Note: This table demonstrates diffusion results in networks of different sizes and with different shock distributions. The red nodes represents seed players, the dark gray nodes are adopters, and the light gray nodes are agents who are in the giant network but have not adopt the new product yet (either because they decide not to adopt or because they are not exposed to the new product).

2.7.2 Players with Instant Decisions

Table 2.2 and 2.3 present a general report of the simulation results for a design where x_i follows a Bernoulli distribution, with probability 0.4 that agent i is a type-0 agent. The two tables demonstrate the same pattern: as the number of observed networks increases, the root mean squared errors of the parameters decrease. In addition, comparing different network sizes when $n = 1000$ and $n = 2000$ in Table 2.2 (or Table 2.3), RMSEs are smaller if the network size is larger, as implied by Theorem 6. Moreover, the distribution of individual payoff noises ε_i affects the estimation accuracy, as the RMSEs are generally smaller for the model with a normally distributed error. Table 2.6 and 2.7 in the appendix present the simulation results from larger networks with 1000 and 2000 nodes and with players who make instant decisions. They demonstrate the bias, variance and RMSE in the vector form.

Table 2.2: Estimating σ and θ when ε is Uniformly Distributed

		σ			θ		
Network Size	# networks	BIAS	VAR	RMSE	BIAS	VAR	RMSE
200	1	0.0370	0.8754	0.8616	0.1671	0.3500	0.5101
	5	0.0148	0.1212	0.1335	0.4251	0.2403	0.6605
	10	0.0126	0.0526	0.0642	0.2216	0.2141	0.4314
500	1	0.0409	0.2359	0.2720	0.1152	0.2372	0.3476
	5	0.0308	0.0389	0.0689	0.1452	0.2348	0.3753
	10	0.0300	0.0132	0.0430	0.1018	0.1840	0.2821
800	1	0.0066	0.1903	0.1931	0.0526	0.1751	0.2241
	5	0.0025	0.0317	0.0336	0.0464	0.1355	0.1792
	10	0.0021	0.0166	0.0183	0.0692	0.1304	0.1969

Note: The bias, standard deviation and root mean squared errors in this table are calculated from S = 50 independent draws of r networks, where each network contains n individuals. The agents make instant adoption decisions when they become informed. $\varepsilon \sim U(0, 1)$).

Table 2.3: Estimating σ and θ when ε is Normally Distributed

		σ			θ		
Network Size	# networks	BIAS	VAR	RMSE	BIAS	VAR	RMSE
200	1	0.0455	0.7218	0.7337	0.0103	0.1014	0.1097
	5	0.0230	0.1049	0.1258	0.0076	0.0152	0.0225
	10	0.0318	0.0449	0.0758	0.0072	0.0090	0.0160
500	1	0.0180	0.2790	0.2915	0.0140	0.0407	0.0538
	5	0.0128	0.0431	0.0550	0.0129	0.0090	0.0218
	10	0.0155	0.0190	0.0341	0.0126	0.0022	0.0147
800	1	0.0083	0.1494	0.1547	0.0063	0.0292	0.0349
	5	0.0011	0.0256	0.0262	0.0007	0.0063	0.0068
	10	0.0009	0.0143	0.0149	0.0005	0.0030	0.0034

Note: The bias, standard deviation and root mean squared errors in this table are calculated from S = 50 independent draws of r networks, where each network contains n individuals. The agents make instant adoption decisions when they become informed. $\varepsilon \sim N(0.5, 0.3)$.

2.7.3 Players with Deferred Decisions

The same simulations are perform on networks with players who are allowed to postpone their adoption decisions. There is an increase in the number of adopters because of this slight change in diffusion mechanism, which resulting in more observations from certain networks.

Table 2.4: Estimating σ and θ when ε is Uniformly Distributed

		σ			θ		
Network Size	# networks	BIAS	VAR	RMSE	BIAS	VAR	RMSE
200	1	0.0157	0.2083	0.2198	0.0841	0.0878	0.1701
	5	0.0105	0.0329	0.0427	0.0345	0.0659	0.0991
	10	0.0121	0.0150	0.0269	0.0560	0.0517	0.1067
500	1	0.0024	0.0951	0.0956	0.1067	0.0557	0.1613
	5	0.0050	0.0136	0.0184	0.0258	0.0353	0.0604
	10	0.0043	0.0050	0.0092	0.0139	0.0199	0.0334
800	1	0.0042	0.0525	0.0556	0.0679	0.0640	0.1307
	5	0.0008	0.0109	0.0115	0.0204	0.0260	0.0459
	10	0.0004	0.0051	0.0055	0.0152	0.0157	0.0306

Note: The bias, standard deviation and root mean squared errors in this table are calculated from S = 50 independent draws of r networks, where each network contains n individuals. The agents are allowed to defer their adoption decisions. $\varepsilon \sim U(0, 1)$).

Table 2.5: Estimating σ and θ when ε is Normally Distributed

		σ			θ		
Network Size	# networks	BIAS	VAR	RMSE	BIAS	VAR	RMSE
200	1	0.0042	0.0525	0.0556	0.0679	0.0640	0.1307
	5	0.0008	0.0109	0.0115	0.0204	0.0260	0.0459
	10	0.0004	0.0051	0.0055	0.0152	0.0157	0.0306
500	1	0.0035	0.0810	0.0829	0.0032	0.0242	0.0269
	5	0.0023	0.0116	0.0137	0.0042	0.0061	0.0102
	10	0.0021	0.0054	0.0074	0.0038	0.0025	0.0062
800	1	0.0020	0.0592	0.0600	0.0041	0.0207	0.0244
	5	0.0005	0.0098	0.0101	0.0018	0.0043	0.0060
	10	0.0003	0.0044	0.0047	0.0020	0.0014	0.0034

Note: The bias, standard deviation and root mean squared errors in this table are calculated from S = 50 independent draws of r networks, where each network contains n individuals. The agents are allowed to defer their adoption decisions. $\varepsilon \sim N(0.5, 0.3)$.

2.8 Conclusion

We develop a structural model to analyze diffusion processes with players who observe their neighbors and form beliefs about their neighbors' decisions in fixed networks. We provide an econometric framework for the diffusion processes and prove that, under certain assumptions, one agent's unique equilibrium belief about another agent's choice depends on the characteristics of both sides. The undirected network structure plays a key role in this case. The awareness that one player's own type could affect her neighbors' payoffs, and that her decision could result in a change of the information transmission path, updates an agent's expectation of her neighbors' choices, resulting in heterogeneous rational expectations among players of different types. We then demonstrate the existence of equilibria and characterize the unique solution to the equilibrium beliefs.

We propose a consistent and tractable two-step m-estimator for individual payoffs. The estimator requires that, for each player i who is informed, researchers observe the types and degrees of player i and of the person who passed the information to player i. The estimator is consistent under the asymptotics that the number of players in a single large network grows to infinity. Then we address the selectivity using a novel approach based on the tree network.

Monte Carlo simulations demonstrate good performance with finite samples. The method can be used in empirical applications like product referrals, technology adoption, rumors on social media, etc.

2.9 A. Discussion

2.9.1 Irreversible Decisions

We assume players make once-and-for-all choices in the diffusion process; whether or not they adopt, they do not revisit their decisions. Recall the example of the diffusion of microfinance in the introduction, where players decide whether to apply for the loan once they learn about it from their neighbors. The adoption decision is irreversible.

We justify this assumption from four perspectives:

a) High switching cost. Although the decision to adopt a new product or a new technology is often reversible in the long run, it might not be the case in the short run due to high switching cost - the cost of getting the new product or the necessary capital investment in the new technology could be prohibitive. For instance, when a student gets a Mac for college, the chance that she will switch to a PC in a short time is small.

In addition, short-run diffusion results are economically important because adoption of new technologies often has been slow (Ryan and Gross 1943[50], Griliches 1957[28], Munshi 2004[45], Skinner and Staiger 2009[54]). Ryan and Gross (1943)[50] show that it took ten years for hybrid seed corn to be adopted in Iowa in the 1930s.

b) Unobserved feedback from neighbor adopters. In some cases, it might be reasonable to think that players would postpone their adoption decisions in order to process information from their neighbors. However, sometimes the feedback is unobservable, and therefore would not give players information that might cause them to revisit their decisions. This is also illustrated in Banerjee et al. (2013)[6]. In their paper, information about microfinance loans spread among villagers. Agents made decisions within one period (4 months), but it took 1 to 2 years to observe the long-run feedback from their neighbors and learn whether neighbors did well with microfinance. Thus, there is little chance that players will revisit their decisions in the long run. The data also show that most adoption happens within one year after the first microfinance loan in the village.

Referral programs are another case in which feedback is usually unobservable. When a person receives referred and has to make a quick adoption decision, it is very probable that the only feedback she can get is from the referrer. The referrer would thus play a significant role in the player's decision-making process.

c) Irreversible adoption of information. The diffusion of information is naturally irreversible. Watching a film, listening to music, and reading a book are all examples of one-time consumption that is not returnable. Another case of information diffusion is knowledge acquisition, such as the adoption of new skills and new technologies.

d) There would be little difference in the results even if players had the option to adopt in later periods. This would be the case even if some players wait until neighbors make their decisions before adopting themselves. Sadler (2020)[51] finds that late adoption would affect the results very little in a sufficiently large random graph. This is because large cascades depend mainly on the group of players who would adopt once they observe one of their neighbors adopt. If it were otherwise, there would be a substantial group of players who would like to wait; in that case, the cascade would not exist at all.

2.9.2 Configuration Model

In network science, the configuration model is a method for generating random networks from a given degree sequence. It is widely used as a reference model for real-life social networks because it allows arbitrary degree distributions. In addition, the multi-type configuration model represents more realistic network features, including heterogeneity and homophily.

A multi-type configuration model can be formed with the following steps: First, take a degree sequence, i.e., assign the number of neighbors of each type X to each player. The degrees of the vertices are represented as half-links or stubs labeled with the type of player to which the vertice connects. The sum of X_i-to-X_j stubs must be equal to the sum of X_j-to-X_i stubs in order to be able to construct a graph. Then choose two stubs uniformly at random and connect them to form an edge. Choose another pair from the remainingstubs and connect them. Continue this step until running out of stubs.

2.10 B. Proofs

Proof of Theorem 4. By equation (2.4), the probability that agent i decide to adopt, given i's neighbors N_i and parent information ω_i, is

$$P(y_i = 1|N_i, \omega_i) = F_\varepsilon\left(h_i + a_i(N_i, \omega_i, \sigma)\right) \tag{2.17}$$

where $a_i(N_i, \omega_i, \sigma) = \gamma_{i,\omega_i} + \sum_{j \in N_i \backslash \omega_i} \gamma_{ij} \cdot \sigma_{ij}$.

Recall that $\tau_l = (x_l, N_l, x_{N_l}, \varepsilon_l, \omega_l)$, by Assumption 4 and the Law of Iterated Expectation,

$$\begin{aligned} \sigma_{li} =& P(y_i = 1|\omega_i = l, \tau_l, l \in \mathcal{I}) \\ =& \sum_{N_i} P(y_i = 1|N_i, \omega_i = l, x_l, N_l, x_{N_l}, \varepsilon_l, \omega_l, l \in \mathcal{I}) \cdot P(N_i|\omega_i = l, x_l, N_l, x_{N_l}, \varepsilon_l, \omega_l, l \in \mathcal{I}). \end{aligned} \tag{2.18}$$

By Assumption 3 that states the independence between ε_l and ε_i,

$$P(y_i = 1|N_i, \omega_i = l, x_l, N_l, x_{N_l}, \varepsilon_l, \omega_l, l \in \mathcal{I}) = P(y_i = 1|N_i, \omega_i = l).$$

With model setup that agent l does not observe the whole network, (N_l, x_{N_l}) contains the same information about agent l's local network as $\mathbf{d}_l$.

$$P(N_i|\omega_i = l, x_l, N_l, x_{N_l}, \varepsilon_l, \omega_l, l \in \mathcal{I}) = P(N_i|\omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}).$$

Thus by equations (2.17) and (2.18),

$$\sigma_{li} = \sum_{N_i} F_\varepsilon\left(h_i + a_i(N_i, \omega_i = l, \sigma)\right) \cdot P(N_i|\omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}),$$

where $a_i(N_i, \omega_i = l, \sigma) = \gamma_{il} + \sum_{j \in N_i \backslash l} \gamma_{ij} \sigma_{ij}$.

Therefore, σ_{li} can be written as a function of σ_{ik} for all $k \neq i$. Denote the projection by $\sigma_{li} = R_{li}\left(\sigma_{i,-i}\right)$. A continuous mapping from $[0,1]^{n(n-1)}$ to $[0,1]^{n(n-1)}$ is defined if the parameters $h(\cdot)$ and $\gamma(\cdot)$ are identified. This leads to the the existence of at least one

Bayesian equilibrium by Brouwer fixed-point theorem.

$\square$

Proof of Proposition 2. First, we show that $\sigma_{ik} = \sigma_{jk}$ if $x_i = x_j, \mathbf{d}_i = \mathbf{d}_j$ for any i, j, k.

$$\sigma_{ik} = \sum_{N_k} F_\varepsilon\left(h_k + a_k(N_k, \omega_k = i, \sigma)\right) \cdot P(N_k|\omega_k = i, x_k, x_i, \mathbf{d}_i, i \in \mathcal{I}),$$
$$\sigma_{jk} = \sum_{N_k} F_\varepsilon\left(h_k + a_k(N_k, \omega_k = j, \sigma)\right) \cdot P(N_k|\omega_k = j, x_k, x_j, \mathbf{d}_j, j \in \mathcal{I}),$$

where

$$a_k(N_k, \omega_k = i, \sigma) = \gamma_{ki} + \sum_{p \in N_k \setminus i} \gamma_{kp}\sigma_{kp} = \gamma(x_i, x_k, \mathbf{d}_k)(1 - \sigma_{ki}) + \sum_{p \in N_k} \gamma(x_p, x_k, \mathbf{d}_k)\sigma_{kp},$$
$$a_k(N_k, \omega_k = j, \sigma) = \gamma_{kj} + \sum_{p \in N_k \setminus j} \gamma_{kp}\sigma_{kp} = \gamma(x_j, x_k, \mathbf{d}_k)(1 - \sigma_{kj}) + \sum_{p \in N_k} \gamma(x_p, x_k, \mathbf{d}_k)\sigma_{kp}.$$

Since $x_i = x_j$ and $\sigma_{ki} = \sigma_{kj}$ by Assumption 6, $a_k(N_k, \omega_k = i, \sigma) = a_k(N_k, \omega_k = j, \sigma)$. Notice that $\sum_{p \in N_k} \gamma(x_p, x_k, \mathbf{d}_k)\sigma_{kp}$ is the same for any k that has the same degree $\mathbf{d}_k$, we can abuse the notation and write $a_k(N_k, \omega_k = i, \sigma) = a_k(\mathbf{d}_k, \omega_k = i, \sigma)$.

Thus

$$\sigma_{ik} = \sum_{\mathbf{d}_k} F_\varepsilon\left(h_k + a_k(\mathbf{d}_k, \omega_k = i, \sigma)\right) \cdot P(\mathbf{d}_k|\omega_k = i, x_k, x_i, \mathbf{d}_i, i \in \mathcal{I}),$$
$$\sigma_{jk} = \sum_{\mathbf{d}_k} F_\varepsilon\left(h_k + a_k(\mathbf{d}_k, \omega_k = j, \sigma)\right) \cdot P(\mathbf{d}_k|\omega_k = j, x_k, x_j, \mathbf{d}_j, j \in \mathcal{I}),$$

In addition, $P(\mathbf{d}_k|\omega_k = i, x_k, x_i, \mathbf{d}_i, i \in \mathcal{I}) = P(\mathbf{d}_k|\omega_k = j, x_k, x_j, \mathbf{d}_j, j \in \mathcal{I})$ holds by the model assumption that agents does not observe the whole network. Thus $\sigma_{ik} = \sigma_{jk}$ if $x_i = x_j, \mathbf{d}_i = \mathbf{d}_j$. Since $\sigma_{ij} = \sigma_{ik}$ if $x_j = x_k$ by Assumption 6, it's easy to show that $\sigma_{ij} = \sigma_{kl}$ if $x_j = x_l, x_i = x_k, \mathbf{d}_i = \mathbf{d}_k$. $\square$

Proof of Theorem 5. Let $\Sigma = [0, 1]^{n^2}$ be the bounded set that contains all possible σ's.

Define a mapping $\boldsymbol{R} : \Sigma \to \Sigma$ such that

$$\boldsymbol{R}(\sigma)_{li} \equiv R_{li}\left(\sigma_{i,-i}\right), \quad \forall\ i \neq l$$

Note that $\sigma_{li} = R_{li}\left(\sigma_{i,-i}\right)$, so $\boldsymbol{R}(\sigma) \in \Sigma$ for any $\sigma \in \Sigma$.

Then we show that $||\boldsymbol{R}(\sigma) - \boldsymbol{R}(\sigma')|| \leq \alpha||\sigma - \sigma'||$ for some $\alpha \in (0,1)$, where $||\cdot||$ is the supreme norm.

First, from the proof of Proposition 2,

$$R_{li}\left(\sigma\right) = \sum_{\mathbf{d}_i} F_{\varepsilon}\left(h_i + a_i(\mathbf{d}_i, \omega_i = l, \sigma)\right) \cdot P(\mathbf{d}_i|\omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}).$$

Notice that in the conditional probability $P(\mathbf{d}_i|\omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I})$, the distribution of $\mathbf{d}_i$ is restricted by the following conditions:

1. one link of i connects to a type-x_l player since $l \in N_i$,
2. $\mathbf{d}_l$ affects the distribution of $\mathbf{d}_i$ because i may be connected to some $j \in N_l$,
3. $l \in \mathcal{I}$ thus $\omega_l \in \mathcal{I}$.

Only the third condition is potentially related to the diffusion process. The first two do not depend on σ since they are determined by the type and degree distributions that govern the network generating process. Sadler (2020)[51] shows that when agent l does not observe the identity of her parent, her beliefs about her neighbors' degree would be affected by the viral belief distortion. However, when ω_l is observed, viral belief distortion would only affect l's belief about ω_l's degree, not other neighbors' degree. In sum, $P(\mathbf{d}_i|\omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I})$ would not change with σ.

Let $N(\mathbf{d}_i)$ denote an arbitrary neighborhood with attribute $\mathbf{d}_i$,

$$a_i = a_i(\mathbf{d}_i, \omega_i = l, \sigma) = \gamma_{il} + \sum_{j \in N(\mathbf{d}_i)\backslash l} \gamma_{ij}\sigma_{ij},$$
$$a_i' = a_i(\mathbf{d}_i, \omega_i = l, \sigma') = \gamma_{il} + \sum_{j \in N(\mathbf{d}_i)\backslash l} \gamma_{ij}\sigma_{ij}',$$

Then for any $\sigma, \sigma' \in \Sigma$,

$$\begin{aligned} |\boldsymbol{R}(\sigma)_{li} - \boldsymbol{R}(\sigma')_{li}| &\leq \sum_{\mathbf{d}_i} |F_\varepsilon(h_i + a_i) - F_\varepsilon(h_i + a_i')| \cdot P(\mathbf{d}_i | \omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}) \\ &\leq \sum_{\mathbf{d}_i} K\,|a_i - a_i'| \cdot P(\mathbf{d}_i | \omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}). \end{aligned}$$

Note that

$$|a_i - a_i'| = \left| \sum_{j \in N(\mathbf{d}_i)\backslash l} \gamma_{ij}\left(\sigma_{ij} - \sigma_{ij}'\right) \right| \leq (\gamma_i \mathbf{d}_i' - \gamma_{il})||\sigma - \sigma'||,$$

where $\gamma_i = (\gamma(X_1, x_i, \mathbf{d}_i), ..., \gamma(X_K, x_i, \mathbf{d}_i))$. Thus

$$\begin{aligned} |\boldsymbol{R}(\sigma)_{li} - \boldsymbol{R}(\sigma')_{li}| &\leq K||\sigma - \sigma'|| \cdot \sum_{\mathbf{d}_i} (\gamma_i \mathbf{d}_i' - \gamma_{il})\mathbf{P}(\mathbf{d}_i | \omega_i = l, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}) \\ &\leq K \cdot \max_{x_i, \mathbf{d}_i} \{\gamma_i \mathbf{d}_i' - \gamma_{i,\min}\} \cdot ||\sigma - \sigma'|| \leq \alpha||\sigma - \sigma'|| \end{aligned}$$

where $\gamma_{i,\min}$ is the minimum element of vector γ_i.

Since $||\boldsymbol{R}(\sigma) - \boldsymbol{R}(\sigma')|| = \sup_{i \neq l}\{|\boldsymbol{R}(\sigma)_{il} - \boldsymbol{R}(\sigma')_{il}|\}$, it follows that $||\boldsymbol{R}(\sigma) - \boldsymbol{R}(\sigma')|| \leq \alpha||\sigma - \sigma'||$. By the contraction mapping theorem, the existence and uniqueness of the solution σ is guaranteed. □

Proof of Proposition 3. By Proposition 2, a unique Bayesian equilibrium exists in each data-generating process indexed by n, and

$$y_i = 1 \quad \text{w.p.} \quad F_\varepsilon\left(h_i + a_i(\mathbf{d}_i, \omega_i, \sigma_n^b)\right).$$

where σ_n^b is the belief, and $a_i(\mathbf{d}_i, \omega_i, \sigma_n^b) = \gamma_{i,\omega_i} + \sum_{N(\mathbf{d}_i)\backslash\omega_i} \gamma_{ij}\sigma_n^b(x_j, x_i, \mathbf{d}_i)$.

Define $\sigma_n(x', x, \mathbf{d}) \equiv \frac{1}{|\mathcal{I}_n|}\sum_{i\neq j} E_n(y_j|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, \omega_j = i, i \in \mathcal{I}_n) = E_n(y_j|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, \omega_j = i, i \in \mathcal{I}_n)$, the second equality comes from the symmetry of beliefs in Proposition 2. Similarly, define

$$\begin{aligned} q_n(\mathbf{d}'|x', x, \mathbf{d}) &\equiv \frac{1}{|\mathcal{I}_n|}\sum_{j\neq i} E_n\left[\mathbb{1}\left\{\mathbf{d}_j = \mathbf{d}'\right\}|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, \omega_j = i, i \in \mathcal{I}_n\right] \\ &= E_n\left[\mathbb{1}\left\{\mathbf{d}_j = \mathbf{d}'\right\}|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, \omega_j = i, i \in \mathcal{I}_n\right]. \end{aligned}$$

Thus by the Law of Iterated Expectation, for any n

$$\begin{aligned} \sigma_n(x', x, \mathbf{d}) = \sum_{\mathbf{d}'} & E_n(y_j|x_j = x', \mathbf{d}_j = \mathbf{d}', x_i = x, \mathbf{d}_i = \mathbf{d}, \omega_j = i, i \in \mathcal{I}_n) \\ & \cdot E_n(\mathbb{1}\left\{\mathbf{d}_j = \mathbf{d}'\right\}|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, \omega_j = i, i \in \mathcal{I}_n) \\ = \sum_{\mathbf{d}'} & F_\varepsilon\left(h(x', \mathbf{d}') + a(x', \mathbf{d}', x_\omega = x, \sigma_n^b)\right) \cdot q_n(\mathbf{d}'|x', x, \mathbf{d}). \end{aligned}$$

where $a(x', \mathbf{d}', x_\omega = x, \sigma_n^b) = \gamma(x, x', \mathbf{d}')(1-\sigma_n^b(x, x', \mathbf{d}')) + \sum_{k=1}^K \gamma(X_k, x', \mathbf{d}')\sigma_n^b(X_k, x', \mathbf{d}')d'_k$.

By Assumption 4, $\sigma_n = \sigma_n^b$. Thus

$$\sigma_n(x', x, \mathbf{d}) = \sum_{\mathbf{d}'} F_\varepsilon\left(h(x', \mathbf{d}') + a(x', \mathbf{d}', x_\omega = x, \sigma_n)\right) \cdot q_n(\mathbf{d}'|x', x, \mathbf{d}).$$

is a self-map over the set of $\{\sigma_n\}$.

Next, we show that σ_n, the solution of the self-map is unique. Suppose there exists another solution $\tilde{\sigma}_n$, then

$$\begin{aligned}|\sigma_{n,li}-\tilde{\sigma}_{n,li}| =&\Bigg|\sum_{\mathbf{d}_i} F_\varepsilon\left(h_i + a(x_i,\mathbf{d}_i,\omega_i,\sigma_n)\right)\cdot q(\mathbf{d}_i|x_i,x_l,\mathbf{d}_l)\\ &-\sum_{\mathbf{d}_i} F_\varepsilon\left(h_i + a(x_i,\mathbf{d}_i,\omega_i,\tilde{\sigma_n})\right)\cdot \tilde{q}(\mathbf{d}_i|x_i,x_l,\mathbf{d}_l)\Bigg|\\ \leq&\Bigg|\sum_{\mathbf{d}_i}\left[F_\varepsilon\left(h_i + a(x_i,\mathbf{d}_i,\omega_i,\sigma)\right) - F_\varepsilon\left(h_i + a(x_i,\mathbf{d}_i,\omega_i,\tilde{\sigma})\right)\right]\cdot q(\mathbf{d}_i|x_i,x_l,\mathbf{d}_l)\Bigg|\\ &+\Bigg|\sum_{\mathbf{d}_i} F_\varepsilon\left(h_i + a(x_i,\mathbf{d}_i,\omega_i,\tilde{\sigma})\right)\cdot\left(q(\mathbf{d}_i|x_i,x_l,\mathbf{d}_l)-\tilde{q}(\mathbf{d}_i|x_i,x_l,\mathbf{d}_l)\right)\Bigg|\\ \leq&\alpha||\sigma-\tilde{\sigma}|| + ||q-\tilde{q}||,\end{aligned}$$

we have

$$||\sigma-\tilde{\sigma}|| \leq \frac{||q-\tilde{q}||}{1-\alpha}.$$

where q and $\tilde{q}$ denote the generic density (probability mass) function of $\mathbf{d}'$ given x', x, m. Therefore for any $\bar{c} > 0$, there exists $c > 0$ such that $||q-\tilde{q}|| \leq c$ implies $||\sigma-\tilde{\sigma}|| \leq \bar{c}$, where σ and $\tilde{\sigma}$ are unique solutions to $\sigma = \boldsymbol{R}(\sigma; q)$ and $\tilde{\sigma} = \boldsymbol{R}(\tilde{\sigma}; \tilde{q})$. As the q^* exists in limit, σ^* also exists in limit. □

Proof of Proposition 4. Recall that

$$S^*(x_{\omega_i}, x_i, \mathbf{d}_i) = h(x_i,\mathbf{d}_i) + T(x_i,\mathbf{d}_i) + \gamma(x_{\omega_i}, x_i, \mathbf{d}_i)\left(1-\sigma^*(x_{\omega_i}, x_i, \mathbf{d}_i)\right)$$

Note that $T(x_i,\mathbf{d}_i) = \sum_{j\in N(\mathbf{d}_i)}\gamma(x_j,x_i,\mathbf{d}_i)\sigma^*(x_j,x_i,\mathbf{d}_i)$ and

$$\gamma(x_j,x_i,\mathbf{d}_i) = \frac{S^*(x_j,x_i,\mathbf{d}_i) - h(x_i,\mathbf{d}_i) - T(x_i,\mathbf{d}_i)}{1-\sigma^*(x_j,x_i,\mathbf{d}_i)}$$

Then

$$T(x_i,\mathbf{d}_i) = \sum_{j\in N(\mathbf{d}_i)}\frac{S^*(x_j,x_i,\mathbf{d}_i) - h(x_i,\mathbf{d}_i) - T(x_i,\mathbf{d}_i)}{1-\sigma^*(x_j,x_i,\mathbf{d}_i)}\sigma^*(x_j,x_i,\mathbf{d}_i)$$

Let $X^*(x_i,\mathbf{d}_i) = \sum_{j\in N(\mathbf{d}_i)} S^*(x_j,x_i,\mathbf{d}_i)\frac{\sigma^*(x_j,x_i,\mathbf{d}_i)}{1-\sigma^*(x_j,x_i,\mathbf{d}_i)}$, $Y^*(x_i,\mathbf{d}_i) = \sum_{j\in N(\mathbf{d}_i)}\frac{\sigma^*(x_j,x_i,\mathbf{d}_i)}{1-\sigma^*(x_j,x_i,\mathbf{d}_i)}$, reshape the equation and we can get $T(x_i, \mathbf{d}_i)\,[1 + Y^*(x_i, \mathbf{d}_i)] = X^*(x_i, \mathbf{d}_i) - h(x_i, \mathbf{d}_i)Y^*(x_i, \mathbf{d}_i)$ or $T = (X^* - hY^*)/(1 + Y^*)$. Replacing $T(x_i, \mathbf{d}_i)$ in $S(x_{\omega_i}, x_i, \mathbf{d}_i)$ with the above expression,

$$S^*(x_j, x_i, \mathbf{d}_i)(1+Y^*(x_i, \mathbf{d}_i)) - X^*(x_i, \mathbf{d}_i) = h(x_i, \mathbf{d}_i) + \gamma(x_j, x_i, \mathbf{d}_i)[(1-\sigma^*(x_j, x_i, \mathbf{d}_i))(1+Y^*(x_i, \mathbf{d}_i))].$$

It is easy to see, with the above equation, that the identification is not achievable even if we identify $S^*(x_j, x_i, \mathbf{d}_i)$ and $\sigma^*(x_j, x_i, \mathbf{d}_i)$ for all combinations of $(x_j, x_i, \mathbf{d}_i)$ because for each $(x_i, \mathbf{d}_i)$, we have K equalities but $K + 1$ unknowns to be identified. Therefore, we need to impose additional assumption to ensure identification.

For the following proof for identification, let $A^*(x_j, x_i, \mathbf{d}_i) = S^*(x_j, x_i, \mathbf{d}_i)(1+Y^*(x_i, \mathbf{d}_i)) - X^*(x_i, \mathbf{d}_i)$, $B^*(x_j, x_i, \mathbf{d}_i) = (1-\sigma^*(x_j, x_i, \mathbf{d}_i))(1+Y^*(x_i, \mathbf{d}_i))$, which are both point-identified. Thus $A^*(x_j, x_i, \mathbf{d}_i) = h(x_i, \mathbf{d}_i) + \gamma(x_j, x_i, \mathbf{d}_i)B^*(x_j, x_i, \mathbf{d}_i)$.

I propose two types of assumptions that ensures identification. Under Assumption 9(a), $A^*(x_j, x_i, \mathbf{d}_i) = h(x_i, \mathbf{d}_i) + \gamma(x_j, x_i)B^*(x_j, x_i, \mathbf{d}_i)$. for each x_i. The identification is achieved when the set of equalities satisfies a rank condition such that for each x, there exist $(x_i, x_j, \mathbf{d}_m, \mathbf{d}_n)$ such that

$$\begin{pmatrix} B^*(x_i, x, \mathbf{d}_m) & B^*(x_j, x, \mathbf{d}_n) \\ B^*(x_i, x, \mathbf{d}_n) & B^*(x_j, x, \mathbf{d}_m) \end{pmatrix}$$

has full rank, or equivalently, $\frac{1-\sigma^*(x_i,x,\mathbf{d}_m)}{1-\sigma^*(x_j,x,\mathbf{d}_m)} \neq \frac{1-\sigma^*(x_i,x,\mathbf{d}_n)}{1-\sigma^*(x_j,x,\mathbf{d}_n)}$.

For $K \geq 2$, under Assumption 9(b), a sufficient condition for identification is that for each x, there exists some type x_i and degree $\mathbf{d}_j$ such that $2A_{ii}A_{jj} + 2A_{ij}A_{ji} > (A_{ii} + A_{jj})(A_{ij} + A_{ji})$, where $A_{ij} = A^*(x_i, x, \mathbf{d}_j)$.

For the case that there is only a single type of agent, $\gamma(d_i) = \phi/d_i$, where d_i is the degree of agent i. ϕ is then identified from $A^*(d) = h + \frac{\phi}{d} \cdot B^*(d)$ since $d \geq 2$.

$\square$

Proof of Proposition 5.

$$\left| \frac{\sum_i y_i \mathbb{1}\{x_i = x'\}\iota_{\omega_i}(x, \mathbf{d})}{\sum_i \mathbb{1}\{x_i = x'\}\iota_{\omega_i}(x, \mathbf{d})} - \sigma^*(x', x, \mathbf{d}) \right| \xrightarrow{p} 0$$

By Proposition 2, the equilibrium beliefs only depends on the agent's and her neighbors' characteristics, not on the identity of agents. Thus for any referrer-receiver pairs with types and degree $(x', x, \mathbf{d})$ (the first argument denotes the type of the receiver, and the second and the third are the type and degree of the referrer), $E(y_i|x_i = x', x_{\omega_i} = x, \mathbf{d}_{\omega_i} = \mathbf{d}) = \sigma(x', x, \mathbf{d})$. As a result,

$$\frac{\sum_{i\in\mathcal{I}_n} E(y_i|x_i, x_{\omega_i}, \mathbf{d}_{\omega_i})\mathbb{1}\{x_i = x'\}\iota_{\omega_i}(x,\mathbf{d})}{\sum_{i\in\mathcal{I}_n} \mathbb{1}\{x_i = x'\}\iota_{\omega_i}(x,\mathbf{d})} = \sigma(x', x, \mathbf{d})$$

Denote the types and degree $(x', x, \mathbf{d})$ by $\tilde{S}$, the support of all types and degree $\{(x', x, \mathbf{d})\}$ by $\mathbf{S}$. Let $\Delta(\tilde{S}) = \frac{\sum_i y_i \mathbb{1}\{x_i=x'\}\iota_{\omega_i}(x,\mathbf{d})}{\sum_i \mathbb{1}\{x_i=x'\}\iota_{\omega_i}(x,\mathbf{d})} - \sigma^*(x', x, \mathbf{d})$ denote the difference. We need to show $\lim_{\eta\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(\sup_{\tilde{S}\in\mathbf{S}} |\Delta(\tilde{S})| > \eta/\sqrt{n}\right) = 0$ for the root-n uniform convergence in probability.

Note that y_i is a function of $x_i, \mathbf{d}_i, \varepsilon_i$. For any i, j such that $x_i = x_j = x'$, $x_{\omega_i} = x_{\omega_j} = x$, and $\mathbf{d}_{\omega_i} = \mathbf{d}_{\omega_j} = \mathbf{d}$, $Cov(y_i, y_j) = E_{\mathbf{d}_i^f(x),\mathbf{d}_j^f(x)}[F_\varepsilon(h_i + \gamma_{i,\omega_i} + \vec{\gamma}_i \mathbf{d}_i^f(x))F_\varepsilon(h_j + \gamma_{j,\omega_j} + \vec{\gamma}_j \mathbf{d}_j^f(x))] - E(y_i)E(y_j) \sim \vec{\gamma}_i' Cov(\mathbf{d}_i^f(x), \mathbf{d}_j^f(x))\vec{\gamma}_j$. Recall that $Cov(\mathbf{d}_i^f(x), \mathbf{d}_j^f(x)) = O_p(n^{-1})$.

Note that $\Delta(\tilde{S}) = \frac{\sum_{i\in\mathcal{I}_n}\left(y_i - E(y_i|x_i,x_{\omega_i},\mathbf{d}_{\omega_i})\right)\mathbb{1}\{x_i=x'\}\iota_{\omega_i}(x,\mathbf{d})}{\sum_{i\in\mathcal{I}_n}\mathbb{1}\{x_i=x'\}\iota_{\omega_i}(x,\mathbf{d})}$, the variance of $\Delta(\tilde{S})$ is

$$\begin{aligned}
E[\Delta^2(\tilde{S})] &= Var\left(\frac{\sum_{i\in\mathcal{I}_n} y_i \mathbb{1}\{(x_i, x_{\omega_i}, \mathbf{d}_{\omega_i}) = \tilde{S}\}}{\sum_{i\in\mathcal{I}_n} \mathbb{1}\{(x_i, x_{\omega_i}, \mathbf{d}_{\omega_i}) = \tilde{S}\}}\right) = \frac{\frac{1}{|\mathcal{I}_n|^2}\sum_{i\in\mathcal{I}_n} Var(y_i)\mathbb{1}\{(x_i, x_{\omega_i}, \mathbf{d}_{\omega_i}) = \tilde{S}\}}{\left(\frac{1}{|\mathcal{I}_n|}\sum_{i\in\mathcal{I}_n} \mathbb{1}\{(x_i, x_{\omega_i}, \mathbf{d}_{\omega_i}) = \tilde{S}\}\right)^2} + O_p(n^{-1}) \\
&\le \frac{1}{4|\mathcal{I}_n|}\left(\frac{1}{|\mathcal{I}_n|}\sum_{i\in\mathcal{I}_n} \mathbb{1}\{(x_i, x_{\omega_i}, \mathbf{d}_{\omega_i}) = \tilde{S}\}\right)^{-1} + O_p(n^{-1}) = O_p(n^{-1})
\end{aligned}$$

The second equality is a result of $Cov(y_i, y_j) = O_p(n^{-1})$. And the last equality is because size of the informed group $\mathcal{I}_n$ is proportional to n, i.e., $|\mathcal{I}_n|/n \to \alpha_0$ for some $\alpha_0 > 0$ by the definition of *large cascades*.

Then for any η,

$$P\left(\sup_{\tilde{S}\in\mathbf{S}} \left|\Delta(\tilde{S})\right| > \frac{\eta}{\sqrt{n}}\right) \le \sum_{\tilde{S}\in\mathbf{S}} P\left(\Delta^2(\tilde{S}) > \frac{\eta^2}{n}\right) \le \sum_{\tilde{S}\in\mathbf{S}} E\left(\Delta^2(\tilde{S})\right) n\eta^{-2} \le |\mathbf{S}| n\eta^{-2} \sup_{\tilde{S}\in\mathbf{S}} E\left(\Delta^2(\tilde{S})\right).$$

Hence, $\lim_{\eta\to\infty}\limsup_{n\to\infty} P\left(\sup_{\tilde{S}\in\mathbf{S}}\left|\Delta(\tilde{S})\right| > \frac{\eta}{\sqrt{n}}\right) = 0$, which establishes the result. $\square$

Proof of Theorem 6. Denote $E(\hat{Q}_{n,k}(\theta,\sigma)|Y)$ by $Q_{n,k}(\theta,\sigma)$. We show in the following proof that there exists an identifiably unique θ_0 that $i)$ maximizes $Q_{n,k}(\theta,\sigma_0)$, and $ii)$ $\sup_\theta|\hat{Q}_{n,k}(\theta,\hat{\sigma})-Q_{n,k}(\theta,\sigma_0)| \xrightarrow{p} 0$, thus $\hat{\theta}-\theta_0 \xrightarrow{p} 0$.

First, we show that the maximizer is unique. for any θ such that $|\theta-\theta_0|\geq\nu>0$,

$$
\begin{aligned}
&\liminf_n Q_{n,k}(\theta_0,\sigma_0) - Q_{n,k}(\theta,\sigma_0)\\
=&\liminf_n E\left(\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} y_i\log\frac{F_\varepsilon\left(Z_i(\sigma_0)'\theta_0\right)}{F_\varepsilon\left(Z_i(\sigma_0)'\theta\right)} + (1-y_i)\log\frac{1-F_\varepsilon\left(Z_i(\sigma_0)'\theta_0\right)}{1-F_\varepsilon\left(Z_i(\sigma_0)'\theta\right)}\middle|Y\right)\\
\geq&\liminf_n \frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} -\log\{F_\varepsilon\left(Z_i(\sigma_0)'\theta\right)+1-F_\varepsilon\left(Z_i(\sigma_0)'\theta\right)\} = 0 \quad a.s.
\end{aligned}
$$

Since $F_\varepsilon(\cdot)$ is strictly increasing, the inequality hold strictly when there exists i such that

$$Z_i(\sigma_0)'(\theta-\theta_0)\neq 0.$$

With condition (b) the claim is proved.

$$\liminf_n \frac{1}{|\mathcal{I}_{n,k}|}(\theta-\theta_0)'Z'Z(\theta-\theta_0) \geq \nu^2 \liminf_n \frac{1}{|\mathcal{I}_{n,k}|}Z'Z > 0.$$

Then we show that $\sup_\theta|\hat{Q}_{n,k}(\theta,\hat{\sigma})-Q_{n,k}(\theta,\sigma_0)| \xrightarrow{p} 0$. Notice that

$$\sup_\theta|\hat{Q}_{n,k}(\theta,\hat{\sigma})-Q_{n,k}(\theta,\sigma_0)| \leq \underbrace{\sup_\theta|\hat{Q}_{n,k}(\theta,\hat{\sigma})-\hat{Q}_{n,k}(\theta,\sigma_0)|}_{A} + \underbrace{\sup_\theta|\hat{Q}_{n,k}(\theta,\sigma_0)-Q_{n,k}(\theta,\sigma_0)|}_{B}.$$

Let

$$f_i(\theta,\sigma) = y_i\log F_\varepsilon\left(Z_i(\sigma)'\theta\right) + (1-y_i)\log(1-F_\varepsilon\left(Z_i(\sigma)'\theta\right)).$$

Notice that

$$||Z_i(\hat{\sigma})-Z_i(\sigma_0)|| \leq (M-1)||\hat{\sigma}-\sigma_0|| \xrightarrow{p} 0,$$

by condition (a) Θ is compact, it follows that $\{f_i(\theta,\cdot)\}_{\theta\in\Theta}$ is equicontinuous. Therefore,

$$A = \sup_{\theta}\left|\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} f_i(\theta,\hat{\sigma}) - f_i(\theta,\sigma_0)\right| \leq \frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} \sup_{\theta}|f_i(\theta,\hat{\sigma}) - f_i(\theta,\sigma_0)| \xrightarrow{p} 0.$$

The claim that B converges in probability to zero follows Theorem 2.1 of Newey (1991). With condition (a) compactness is satisfied. The sequence of nonrandom functions $\{Q_{n,k}(\cdot,\sigma_0)\}$ with $Q_{n,k}(\theta,\sigma_0)$ equal to

$$\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} F_\varepsilon\left(Z_i'\theta\right)\log F_\varepsilon\left(Z_i'\theta\right) + (1 - F_\varepsilon\left(Z_i'\theta\right))\log(1 - F_\varepsilon\left(Z_i'\theta\right)).$$

is equicontinuous as $\{Z_i'\theta\}$ is bounded.

Then we show that the sequence $\{\hat{Q}_{n,k}(\cdot,\sigma_0)\}$ is asymptotically uniformly equicontinuous, i.e.,

$$\limsup_{n\to\infty} P\left(\sup_{||\theta-\theta'||\leq\delta}|\hat{Q}_{n,k}(\theta,\sigma_0) - \hat{Q}_{n,k}(\theta',\sigma_0)| \geq \varepsilon\right) \xrightarrow{\delta\to 0} 0.$$

It suffices to show that

$$\limsup_{n\to\infty} E\left(\sup_{||\theta-\theta'||\leq\delta}|\hat{Q}_{n,k}(\theta,\sigma_0) - \hat{Q}_{n,k}(\theta',\sigma_0)|\right) \xrightarrow{\delta\to 0} 0. \tag{2.19}$$

Notice that $|\hat{Q}_{n,k}(\theta,\sigma_0) - \hat{Q}_{n,k}(\theta',\sigma_0)|$

$$=\left|\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} y_i(\log F_\varepsilon\left(Z_i'\theta\right) - \log F_\varepsilon\left(Z_i'\theta'\right)) + (1-y_i)(\log(1 - F_\varepsilon\left(Z_i'\theta\right)) - \log(1 - F_\varepsilon\left(Z_i'\theta'\right)))\right|$$
$$\leq\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} \max\left\{\left|\log F_\varepsilon\left(Z_i'\theta\right) - \log F_\varepsilon\left(Z_i'\theta'\right)\right|, \left|\log(1 - F_\varepsilon\left(Z_i'\theta\right)) - \log(1 - F_\varepsilon\left(Z_i'\theta'\right))\right|\right\}.$$

Since $\log(F_\varepsilon(Z_i'\cdot))$ and $\log(1-F_\varepsilon(Z_i'\cdot))$ are continuous with bounded differentials, $|\hat{Q}_{n,k}(\theta,\sigma_0) - \hat{Q}_{n,k}(\theta',\sigma_0)| \leq C||\theta-\theta'||$, and equation (2.19) follows.

The last condition to show is pointwise convergence, that is,

$$\hat{Q}_{n,k}(\theta,\sigma_0) - Q_{n,k}(\theta,\sigma_0) \xrightarrow{p} 0, \quad \text{for all} \ \ \theta \in \Theta.$$

Since

$$\hat{Q}_{n,k}(\theta,\sigma_0) - Q_{n,k}(\theta,\sigma_0) = \frac{1}{|\mathcal{I}_{n,k}|} \sum_{i \in \mathcal{I}_{n,k}} [y_i - F_\varepsilon(Z_i(\sigma_0)'\theta)] \log \frac{F_\varepsilon\left(Z_i(\sigma_0)'\theta\right)}{1 - F_\varepsilon\left(Z_i(\sigma_0)'\theta\right)},$$

Let $g_i = [y_i - F_\varepsilon(Z_i(\sigma_0)'\theta)] \log \frac{F_\varepsilon(Z_i(\sigma_0)'\theta)}{1-F_\varepsilon(Z_i(\sigma_0)'\theta)}$. Since the outcomes $\{y_i\}$ are conditionally independent, $\{g_i\}$ are conditionally independent with mean zero and uniformly bounded variance. By the same logic of the proof of Proposition 5, pointwise convergence is established. Therefore $\sup_\theta |\hat{Q}_{n,k}(\theta,\sigma_0) - Q_{n,k}(\theta,\sigma_0)| \xrightarrow{p} 0$ follows as a result of Newey (1991). $\square$

Proof of Theorem 7. Define $\hat{g}_{n,k}(\theta,\sigma) = \nabla_\theta \hat{Q}_{n,k}(\theta,\sigma)$, $g_{n,k}(\theta,\sigma) = E[\nabla_\theta \hat{Q}_{n,k}(\theta,\sigma)|Y]$. Thus

$$\hat{g}_{n,k}(\theta,\sigma) = \frac{1}{|\mathcal{I}_{n,k}|} \sum_{i \in \mathcal{I}_{n,k}} \nabla_\theta f_i(\theta,\sigma)$$

where $f_i(\theta,\sigma) = y_i \log F_\varepsilon\left(Z_i(\sigma)'\theta\right) + (1-y_i)\log(1 - F_\varepsilon\left(Z_i(\sigma)'\theta\right))$.

Since $\hat{\theta}$ solves $\hat{g}_{n,k}(\hat{\theta},\hat{\sigma}) = 0$, by the mean-value theorem,

$$0 = \hat{g}_{n,k}(\theta_0,\hat{\sigma}) + \nabla_\theta \hat{g}_{n,k}(\tilde{\theta},\hat{\sigma})(\hat{\theta} - \theta_0),$$

where $\tilde{\theta}$ is in between $\hat{\theta}$ and θ_0. Then

$$\sqrt{|\mathcal{I}_{n,k}|}(\hat{\theta} - \theta_0) = -\left(\frac{1}{|\mathcal{I}_{n,k}|} \sum_{i \in \mathcal{I}_{n,k}} \nabla_{\theta\theta} f_i(\tilde{\theta},\hat{\sigma})\right)^{-1} \left(\frac{1}{\sqrt{|\mathcal{I}_{n,k}|}} \sum_{i \in \mathcal{I}_{n,k}} \nabla_\theta f_i(\theta_0,\hat{\sigma})\right). \tag{2.20}$$

First, we show that

$$\frac{1}{|\mathcal{I}_{n,k}|} \sum_{i \in \mathcal{I}_{n,k}} \nabla_{\theta\theta} f_i(\tilde{\theta},\hat{\sigma}) - E\left[\frac{1}{|\mathcal{I}_{n,k}|} \sum_{i \in \mathcal{I}_{n,k}} \nabla_{\theta\theta} f_i(\theta_0,\sigma_0) \middle| Y\right] \xrightarrow{p} 0. \tag{2.21}$$

It holds because left-hand side is equal to

$$\underbrace{\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}}\nabla_{\theta\theta}f_i(\tilde{\theta},\hat{\sigma})-\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}}\nabla_{\theta\theta}f_i(\theta_0,\sigma_0)}_{A}$$

$$+\underbrace{\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}}\nabla_{\theta\theta}f_i(\theta_0,\sigma_0)-E\left[\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}}\nabla_{\theta\theta}f_i(\theta_0,\sigma_0)\middle|Y\right]}_{B}$$

Notice that

$$A\leq\sup_{Y}|\nabla_{\theta\theta}f_i(\tilde{\theta},\hat{\sigma})-\nabla_{\theta\theta}f_i(\theta_0,\sigma_0)|$$

A is $o_p(1)$ since $\tilde{\theta}-\theta_0\xrightarrow{p}0$, $\hat{\sigma}-\sigma_0\xrightarrow{p}0$, and $\nabla_{\theta\theta}f_i(\cdot,\cdot)$ is continuous in θ and σ.

B is also $o_p(1)$ since B is equal to

$$\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}}(y_i-F_\varepsilon(a_i))\left(\frac{f_\varepsilon(a_i)'F_\varepsilon(a_i)-f_\varepsilon(a_i)^2}{F_\varepsilon(a_i)^2}+\frac{f_\varepsilon(a_i)'(1-F_\varepsilon(a_i))+f_\varepsilon(a_i)^2}{(1-F_\varepsilon(a_i))^2}\right)Z_i(\sigma_0)Z_i(\sigma_0)'$$

where $a_i=Z_i(\sigma_0)'\theta_0$. Therefore B is $o_p(1)$ since each term in the summation is conditionally independent, mean zero and bounded in variance.

Define $\Gamma_{n,k}=\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}}E\left[\nabla_{\theta\theta}f_i(\theta_0,\sigma_0)\middle|Y\right]$. By equation (2.20) and (2.21),

$$\sqrt{|\mathcal{I}_{n,k}|}(\hat{\theta}-\theta_0)=-\left(\Gamma_{n,k}+o_p(1)\right)^{-1}\left(\frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}}\nabla_\theta f_i(\theta_0,\hat{\sigma})\right). \tag{2.22}$$

It can be proved that the second part of equation (2.22) converges in probability to

$$\frac{1}{\sqrt{|\mathcal{I}_{n,k}|}}\sum_{i\in\mathcal{I}_{n,k}}\nabla_\theta f_i(\theta_0,\sigma_0)+\frac{1}{\sqrt{|\mathcal{I}_{n,k}|}}\sum_{i\in\mathcal{I}_{n,k}}W_i(\hat{\sigma}-\sigma_0)$$

where W_i is defined later.

The first step is linearization. By Taylor's expansion,

$$\nabla_\theta f_i(\theta_0, \sigma) - \nabla_\theta f_i(\theta_0, \sigma_0) = \nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0)(\sigma - \sigma_0) + (\sigma - \sigma_0)' \nabla_{\sigma\sigma} \nabla_\theta f_i(\theta_0, \tilde{\sigma})(\sigma - \sigma_0)$$

Rearrange the formula,

$$\begin{aligned} & ||\nabla_\theta f_i(\theta_0, \sigma) - \nabla_\theta f_i(\theta_0, \sigma_0) - \nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0)(\sigma - \sigma_0)|| \\ =& (\sigma - \sigma_0)' \nabla_{\sigma\sigma} \nabla_\theta f_i(\theta_0, \tilde{\sigma})(\sigma - \sigma_0) \\ \leq& b(Z_i) ||\sigma - \sigma_0||^2 \end{aligned}$$

Note that $||\hat{\sigma} - \sigma_0|| = O_p(1/\sqrt{n})$, thus

$$\begin{aligned} \frac{1}{\sqrt{|\mathcal{I}_{n,k}|}} \sum_{i \in \mathcal{I}_{n,k}} \nabla_\theta f_i(\theta_0, \hat{\sigma}) = & \frac{1}{\sqrt{|\mathcal{I}_{n,k}|}} \sum_{i \in \mathcal{I}_{n,k}} \nabla_\theta f_i(\theta_0, \sigma_0) \\ & + \frac{1}{\sqrt{|\mathcal{I}_{n,k}|}} \sum_{i \in \mathcal{I}_{n,k}} \nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0)(\hat{\sigma} - \sigma_0) + o_p(1) \end{aligned} \tag{2.23}$$

Secondly, we show that

$$\frac{1}{\sqrt{|\mathcal{I}_{n,k}|}} \sum_{i \in \mathcal{I}_{n,k}} \nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0)(\hat{\sigma} - \sigma_0) - \frac{1}{\sqrt{|\mathcal{I}_{n,k}|}} \sum_{i \in \mathcal{I}_{n,k}} E\left[\nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0) | Y\right](\hat{\sigma} - \sigma_0) \xrightarrow{p} 0.$$

Since

$$\begin{aligned} \Delta_i =& \nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0) - E\left[\nabla_\sigma \nabla_\theta f_i(\theta_0, \sigma_0) | Y\right] \\ =& (y_i - F_{\varepsilon,i}) \left[\nabla_\sigma \left(\frac{f_{\varepsilon,i}}{F_{\varepsilon,i}} \cdot Z_i(\sigma_0) \right) + \nabla_\sigma \left(\frac{f_{\varepsilon,i}}{1 - F_{\varepsilon,i}} \cdot Z_i(\sigma_0) \right) \right] \end{aligned}$$

where $F_{\varepsilon,i} = F_\varepsilon(Z_i(\sigma_0)'\theta_0)$, $f_\varepsilon = f_\varepsilon(Z_i(\sigma_0)'\theta_0)$. Δ_i has zero mean and bounded variance. Combined with conditionally independence,

$$\frac{1}{\sqrt{|\mathcal{I}_{n,k}|}} \sum_{i \in \mathcal{I}_{n,k}} \Delta_i (\hat{\sigma} - \sigma_0) \xrightarrow{p} 0$$

Rewrite equation (2.23),

$$\frac{1}{\sqrt{|\mathcal{I}_{n,k}|}}\sum_{i\in\mathcal{I}_{n,k}}\nabla_\theta f_i(\theta_0,\hat{\sigma}) = \frac{1}{\sqrt{|\mathcal{I}_{n,k}|}}\sum_{i\in\mathcal{I}_{n,k}}\nabla_\theta f_i(\theta_0,\sigma_0) + \frac{1}{\sqrt{|\mathcal{I}_{n,k}|}}\sum_{i\in\mathcal{I}_{n,k}} W_i(\hat{\sigma}-\sigma_0) + o_p(1) \tag{2.24}$$

where $W_i = E\left[\nabla_\sigma \nabla_\theta f_i(\theta_0,\sigma_0)|Y\right]$.

$\{\hat{\sigma}_{ij}\}$ is independent of $\{y_i\}$

$$\begin{aligned}\Omega_{n,k} &\triangleq Var\left(\frac{1}{\sqrt{|\mathcal{I}_{n,k}|}}\sum_{i\in\mathcal{I}_{n,k}}\nabla_\theta f_i(\theta_0,\hat{\sigma})\right)\\ &= \frac{1}{|\mathcal{I}_{n,k}|}\sum_{i\in\mathcal{I}_{n,k}} Var\left(\nabla_\theta f_i(\theta_0,\sigma_0)\right) + \frac{1}{|\mathcal{I}_{n,k}|}\left(\sum_{i\in\mathcal{I}_{n,k}} W_i\right) Var(\hat{\sigma})\left(\sum_{i\in\mathcal{I}_{n,k}} W_i\right)^\top \end{aligned} \tag{2.25}$$

Combine equation (2.22) and (2.25),

$$\sqrt{|\mathcal{I}_{n,k}|}\ \Sigma_{n,k}^{-1}(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0, I_{K+1}) \tag{2.26}$$

where $\Sigma_{n,k} = \Gamma_{n,k}^{-1}\Omega_{n,k}\Gamma_{n,k}^{-1}$.

□

The following are proofs of theorems and propositions in the deferred-decision model extension.

Proof of Theorem 8. The probability of agent i adopting the behavior if her neighbor l adopts follows

$$Pr(y_i = 1|y_l = 1, N_i, x_{N_i}, \mathcal{I}) = E\left[F_\varepsilon\left(h_i + a_i(\sigma)\right)|y_l = 1, N_i, x_{N_i}, \mathcal{I}\right]$$

where $a_i(\sigma) = \gamma_{il} + \sum_{j \in N_i \backslash l} \gamma_{ij} \cdot \max\{y_j, \sigma_{ij}\}$. Therefore

$$Pr(y_i = 1 | y_l = 1, N_i, x_{N_i}, \mathcal{I}) = \sum_{y_{N_i}} F_\varepsilon \left(h_i + \gamma_{il} + \sum_{\substack{k \in N_i \\ j \neq l}} \gamma_{ij} \cdot \max\{y_j, \sigma_{ij}\} \right) Pr(y_{N_i} | y_l = 1, N_i, x_{N_i}, \mathcal{I}) \tag{2.27}$$

where $Pr(y_{N_i} | y_l = 1, N_i, x_{N_i}, \mathcal{I}) = \prod_{j \in N_i, j \neq l} (\sigma_{ij})^{y_j} (1 - \sigma_{ij})^{1 - y_j}$. Recall that

$$\sigma_{li} = Pr(y_i = 1 | y_l = 1, g_{il} = 1, \tau_l, \mathcal{I}) = \sum_{N_i} Pr(y_i = 1 | y_l = 1, N_i, x_{N_i}, \mathcal{I}) Pr(N_i | y_l = 1, g_{il} = 1, \tau_l, \mathcal{I}) \tag{2.28}$$

By equations (2.27) and (2.28), σ_{li} can be written as a function of $\sigma_{i,-i}$ for any $i \neq l$. This leads to the proof of the existence of equilibrium using Brouwer's fixed-point theorem.

For simplicity purpose, I illustrate the detailed proof under the following assumption:

Assumption 15. *Assume $\varepsilon_i \sim U[-1, 1]$, i.e., $F_\varepsilon(x) = \kappa((x + 1)/2)$, where $\kappa(x) = \max\{\min\{x, 1\}, 0\}$. $J(x_i) \leq \alpha/(M - 1)$ for some $\alpha \in (0, 1)$ so that the network effect is small enough. In addition, h_i may not be too large or too small so that neighbors' choices affect individual payoffs. In this paper I assume $h_i \in (-1, 1 - \alpha)$ to maintain linearity of the cdf.*

With uniformly distributed shocks, the probability of agent i adopting the behavior if her neighbor l adopts follows

$$\begin{aligned} Pr(y_i = 1 | y_l = 1, N_i, x_{N_i}, \mathcal{I}) =& E\left[\frac{1}{2}(h_i + a_i(\sigma) + 1) | y_l = 1, N_i, x_{N_i}, \mathcal{I}\right] \\ =& \frac{1}{2}(h_i + E(a_i(\sigma) | y_l = 1, N_i, x_{N_i}, \mathcal{I}) + 1) \end{aligned} \tag{2.29}$$

with

$$E(a_i(\sigma) | y_l = 1, N_i, x_{N_i}, \mathcal{I}) = \gamma_{il} + \sum_{j \in N_i \backslash l} \gamma_{ij} \cdot E[\max\{y_j, \sigma_{ij}\}] = \gamma_{il} + \sum_{j \in N_i \backslash l} \gamma_{ij} z_{ij} \tag{2.30}$$

where $z_{ij} = \sigma_{ij}(2 - \sigma_{ij})$.

By the set of equations (2.28), (2.29) and (2.30), I denote the projection by $\sigma_{li} = R_{li}(\sigma_{i,-i})$ for all $i \neq l$, a continuous mapping from $[-1,1]^{n^2}$ to $[-1,1]^{n^2}$. By Brouwer fixed-point theorem, there is at least one fixed point solution, which gives at least one equilibrium $\{\sigma_{ij}\}$ for $i = 1, ..., n$, $j = 1, ..., n$.

$\square$

Proof of Theorem 9. Let $\mathcal{S} = [-1,1]^{n^2}$ be the bounded set that contains all possible σ's. Define a mapping $\boldsymbol{R} : \mathcal{S} \to \mathcal{S}$ such that

$$\boldsymbol{R}(\sigma)_{li} \equiv R_{li}(\sigma_{i,-i}), \quad \forall i \neq l$$

Note that $\sigma_{li} = R_{li}(\sigma_{i,-i})$, so $\boldsymbol{R}(\sigma) \in \mathcal{S}$ for any $\sigma \in \mathcal{S}$.

By Theorem 8 and Proposition 6, under Assumptions 3, 5, 6, and 11, there exists an equilibrium belief σ^* that satisfies the following self-projection $\sigma = \boldsymbol{R}(\sigma)$,

$$\begin{aligned}\sigma_{li} &= \sum_{N_i} E_Y\left[F_\varepsilon\left(h_i + a_i(\sigma)\right)\right] \cdot P(N_i | y_l = 1, g_{li} = 1, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}) \\ &= \sum_{\mathbf{d}_i} E_Y\left[F_\varepsilon\left(h(x_i, \mathbf{d}_i) + a_i(x_i, \mathbf{d}_i, \sigma)\right)\right] \cdot P(\mathbf{d}_i | y_l = 1, g_{li} = 1, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I})\end{aligned}$$

$\forall l \neq i$, where $a_i(x_i, \mathbf{d}_i, \sigma) = \gamma(x_l, x_i, \mathbf{d}_i) + \sum_{j \in N_i \setminus l} \gamma(x_l, x_i, \mathbf{d}_i) \cdot \max\{y_j, \sigma(x_l, x_i, \mathbf{d}_i)\}$.

Then I show that $||\boldsymbol{R}(\sigma) - \boldsymbol{R}(\sigma')|| \leq \alpha ||\sigma - \sigma'||$ for some $\alpha \in (0,1)$, where $||\cdot||$ is the supreme norm.

For any $\sigma, \sigma' \in \mathcal{S}$, we have

$$
\begin{aligned}
|\boldsymbol{R}(\sigma)_{li} - \boldsymbol{R}(\sigma')_{li}| =& \sum_{\mathbf{d}_i} \left(E_Y \left[F_\varepsilon \left(h(x_i, \mathbf{d}_i) + a_i(x_i, \mathbf{d}_i, \sigma) \right) \right] - E_Y \left[F_\varepsilon \left(h(x_i, \mathbf{d}_i) + a_i(x_i, \mathbf{d}_i, \sigma') \right) \right] \right) \\
& \cdot P(\mathbf{d}_i | y_l = 1, g_{li} = 1, x_i, x_l, \mathbf{d}_l, l \in \mathcal{I}) \\
=& \left| \sum_{\mathbf{d}_i} \frac{1}{2} \left[\sum_j g_{ij} \gamma_{ij} \left(z_{ij} - z_{i'j} \right) - \left(z_{il} - z_{i'l} \right) \right] Pr(\mathbf{d}_i | G_{li} = 1, \mathcal{I}) \right| \\
\leq& \frac{1}{2} \sum_{\mathbf{d}_i} \left[\sum_{j \neq l} g_{ij} \gamma_{ij} \left| z_{ij} - z_{i'j} \right| \right] Pr(\mathbf{d} | G_{li} = 1, \mathcal{I}) \\
\leq& K \bar{\phi} \max_{i \neq j} \left| \sigma_{ij} - \sigma_{i'j} \right| \\
\leq& \alpha || \sigma - \sigma' ||
\end{aligned}
$$

where $z_{ij} = \sigma_{ij} \left(2 - \sigma_{ij} \right)$, $z_{i'j} = \sigma_{i'j} \left(2 - \sigma_{i'j} \right)$, and $\left| z_{ij} - z_{i'j} \right| = \left| \left(\sigma_{ij} - \sigma_{i'j} \right) \left(2 - \sigma_{ij} - \sigma_{i'j} \right) \right| \leq 2 \left| \sigma_{ij} - \sigma_{i'j} \right|$.

Since $||\boldsymbol{R}(\sigma) - \boldsymbol{R}(\sigma')|| = \sup_{i \neq l} \left\{ |\boldsymbol{R}(\sigma)_{li} - \boldsymbol{R}(\sigma')_{li}| \right\}$, it follows that $||\boldsymbol{R}(\sigma) - \boldsymbol{R}(\sigma')|| \leq \alpha || \sigma - \sigma' ||$. By contraction mapping theorem, the uniqueness of the solution σ is guaranteed.

□

Proof of Proposition 7. By Proposition 6, a unique Bayesian equilibrium exists in each data-generating process indexed by n, and

$$
y_i = 1 \quad \text{w.p.} \quad F_\varepsilon \left(h_i + a_i(\mathbf{d}_i, y_{N_i}, \sigma_n^b) \right).
$$

where σ_n^b is the belief, and $a_i(\mathbf{d}_i, y_{N_i}, \sigma_n^b) = \sum_{j \in N_i} \gamma_{ij} (y_j + (1 - y_j) \sigma_n^b(x_j, x_i, \mathbf{d}_i))$.

Define $\sigma_n(x', x, \mathbf{d}) \equiv \frac{1}{|\mathcal{I}_n^2|} \sum_{i \neq j} E_n(y_j | x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1, i \in \mathcal{I}_n) = E_n(y_j | x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1, i \in \mathcal{I}_n)$, the second equality comes from the symmetry of beliefs in Proposition 6. Similarly, define

$$
\begin{aligned}
q_n(\mathbf{d}'|x', x, \mathbf{d}) &\equiv \frac{1}{|\mathcal{I}_n^2|}\sum_{j\neq i} E_n\left[\mathbb{1}\left\{\mathbf{d}_j = \mathbf{d}'\right\}|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1, i \in \mathcal{I}_n\right] \\
&= E_n\left[\mathbb{1}\left\{\mathbf{d}_j = \mathbf{d}'\right\}|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1, i \in \mathcal{I}_n\right].
\end{aligned}
$$

Thus by the Law of Iterated Expectation, for any n

$$
\begin{aligned}
\sigma_n(x', x, \mathbf{d}) = \sum_{\mathbf{d}'} & E_n(y_j|x_j = x', \mathbf{d}_j = \mathbf{d}', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1, i \in \mathcal{I}_n) \\
& \cdot E_n(\mathbb{1}\left\{\mathbf{d}_j = \mathbf{d}'\right\}|x_j = x', x_i = x, \mathbf{d}_i = \mathbf{d}, g_{ij} = 1, i \in \mathcal{I}_n) \\
= \sum_{\mathbf{d}'} & E_{y_N}\left[F_\varepsilon\left(h(x', \mathbf{d}') + a(\mathbf{d}', y_N, \sigma_n^b)\right)\right] \cdot q_n(\mathbf{d}'|x', x, \mathbf{d}).
\end{aligned}
$$

By Assumption 11, $\sigma_n = \sigma_n^b$. Thus

$$
\sigma_n(x', x, \mathbf{d}) = \sum_{\mathbf{d}'} E_{y_N}\left[F_\varepsilon\left(h(x', \mathbf{d}') + a(\mathbf{d}', y_N, \sigma_n)\right)\right] \cdot q_n(\mathbf{d}'|x', x, \mathbf{d}).
$$

is a self-map over the set of $\{\sigma_n\}$.

Next, we show that σ_n, the solution of the self-map is unique. Suppose there exists another solution $\tilde{\sigma}_n$, then

$$
\begin{aligned}
|\sigma_{n,li} - \tilde{\sigma}_{n,li}| =& \Bigg| \sum_{\mathbf{d}_i} E_{y_{N_i}}\left[F_\varepsilon\left(h_i + a(\mathbf{d}_\mathbf{i}', y_{N_i}, \sigma_n)\right)\right] \cdot q(\mathbf{d}_i|x_i, x_l, \mathbf{d}_l) \\
& - \sum_{\mathbf{d}_i} E_{y_{N_i}}\left[F_\varepsilon\left(h_i + a(\mathbf{d}_\mathbf{i}', y_{N_i}, \tilde{\sigma_n})\right)\right] \cdot \tilde{q}(\mathbf{d}_i|x_i, x_l, \mathbf{d}_l) \Bigg| \\
\leq& \Bigg| \sum_{\mathbf{d}_i} \Big[E_{y_{N_i}}\left[F_\varepsilon\left(h_i + a(\mathbf{d}_\mathbf{i}', y_{N_i}, \sigma)\right)\right] - E_{y_{N_i}}\left[F_\varepsilon\left(h_i + a(\mathbf{d}_\mathbf{i}', y_{N_i}, \tilde{\sigma})\right)\right]\Big] \cdot q(\mathbf{d}_i|x_i, x_l, \mathbf{d}_l) \Bigg| \\
& + \Bigg| \sum_{\mathbf{d}_i} E_{y_{N_i}}\left[F_\varepsilon\left(h_i + a(\mathbf{d}_\mathbf{i}', y_{N_i}, \tilde{\sigma})\right)\right] \cdot \left(q(\mathbf{d}_i|x_i, x_l, \mathbf{d}_l) - \tilde{q}(\mathbf{d}_i|x_i, x_l, \mathbf{d}_l)\right) \Bigg| \\
\leq& \alpha||\sigma - \tilde{\sigma}|| + ||q - \tilde{q}||,
\end{aligned}
$$

we have

$$||\sigma - \tilde{\sigma}|| \leq \frac{||q - \tilde{q}||}{1 - \alpha}.$$

where q and $\tilde{q}$ denote the generic density (probability mass) function of $\mathbf{d}'$ given x', x, m. Therefore for any $\bar{c} > 0$, there exists $c > 0$ such that $||q - \tilde{q}|| \leq c$ implies $||\sigma - \tilde{\sigma}|| \leq \bar{c}$, where σ and $\tilde{\sigma}$ are unique solutions to $\sigma = \boldsymbol{R}(\sigma; q)$ and $\tilde{\sigma} = \boldsymbol{R}(\tilde{\sigma}; \tilde{q})$. As the q^* exists in limit, σ^* also exists in limit. $\square$

Proof of Proposition 8. Conditional on the degree sequence $\mathbf{d}_i$, player l's belief about player i's adoption decision is

$$Pr(y_i = 1 | y_l = 1, \mathbf{d}_i, \sigma) = \frac{1}{2}\left(h_i + \gamma_{il}(1 - z_{il}) + \sum_{j \in N_i} \gamma_{ij} z_{ij} + 1\right) \tag{2.31}$$

We are interested in identifying the parameters h and γ for all types.

Under the assumption that $\mathbf{d} \cdot \mathbb{1} \leq M$, the cardinality of set $\{\mathbf{d}_i\}_{i=1}^n$ is finite when $n \to \infty$. Therefore one can take advantage of the large network and observe $Pr(y_i = 1 | y_l = 1, \mathbf{d}_i, \sigma)$ from the data.

Given equation (2.31), suppose there exist $(h, J, \sigma(h, J))$ and $(\tilde{h}, \tilde{J}, \tilde{\sigma}(\tilde{h}, \tilde{J}))$ such that

$$\begin{aligned} Pr(y_i = 1 | y_l = 1, \mathbf{d}_i, \sigma) =& \frac{1}{2}\left(h_i + \gamma_{il}(1 - z_{il}) + \sum_{j \in N_i} \gamma_{ij} z_{ij} + 1\right) \\ =& \frac{1}{2}\left(\tilde{h}_i + \tilde{\gamma}_{il}(1 - z_{il}) + \sum_{j \in N_i} \tilde{\gamma}_{ij} z_{ij} + 1\right) \end{aligned} \tag{2.32}$$

Reshaping equation (2.32),

$$h_i + \gamma_{il}(1 - z_{il}) - \tilde{h}_i - \tilde{\gamma}_{il}(1 - z_{il}) = \sum_{j \in N_i} (\tilde{\gamma}_{ij} - \gamma_{ij}) z_{ij} \tag{2.33}$$

Equation (2.33) indicates that, since $\mathbf{d}_i$ is a random vector,

$$\begin{aligned}\gamma_{il} z_{il} =& \tilde{\gamma}_{il} \tilde{z}_{il} \quad \text{for } i, l \\ h_i + \gamma_{il} =& \tilde{h}_i + \tilde{\gamma}_{il}\end{aligned} \tag{2.34}$$

Recall that

$$\begin{aligned}\sigma_{ji} &= \frac{1}{2}\left(h_i + J_i\left(1 - z_{ij}\right) + 1\right) + \frac{J_i}{2} E\left[\mathbf{d}_i | g_{ij} = 1, \mathcal{I}\right] \cdot z_i \\ &= \frac{1}{2}\left(\tilde{h}_i + \tilde{J}_i\left(1 - \tilde{z}_{ij}\right) + 1\right) + \frac{\tilde{J}_i}{2} E\left[\mathbf{d}_i | g_{ij} = 1, \mathcal{I}\right] \cdot \tilde{z}_i \\ &= \tilde{\sigma}_{ji}\end{aligned} \tag{2.35}$$

It follows that $h_i = \tilde{h}_i$, $\gamma_{ij} = \tilde{\gamma}_{ij}$. □

Proof of Proposition 9. Let $\mathbb{1}_j$ denote $\mathbb{1}\{x_j = x\}$, $\mathbb{1}'_j = \mathbb{1}\{x_j = x'\}$, $\iota_i = \iota_i(x, \mathbf{d})$. Let $\xi_{ij} = (y_j \mathbb{1}'_j \iota_i g_{ij} + y_i \mathbb{1}'_i \iota_j g_{ji})/2$. Then $\xi_{ij} = \xi_{ji}$ and

$$\frac{1}{n(n-1)} \sum_{j \neq i} y_j \mathbb{1}'_j \iota_i g_{ij} = \frac{2}{n(n-1)} \sum_{j > i} \xi_{ij}.$$

By Chebyshev's inequality,

$$P_n\left\{\left|\frac{2}{n(n-1)} \sum_{j>i} \xi_{ij} - E_n\left(\frac{2}{n(n-1)} \sum_{j>i} \xi_{ij}\right)\right| \geq C\right\} \leq \frac{1}{C^2} V_n\left(\frac{2}{n(n-1)} \sum_{j>i} \xi_{ij}\right)$$

where

$$\begin{aligned}V_n\left(\frac{2}{n(n-1)} \sum_{j>i} \xi_{ij}\right) &= \frac{4}{n^2(n-1)^2} \sum_{j>i} \sum_{t>l} C_n(\xi_{ij}, \xi_{tl}) \\ &= \frac{4}{n^2(n-1)^2}\left\{\underbrace{\sum_{j>i} Vn(\xi_{ij})}_{A_1} + \underbrace{\sum_{\substack{j>i \\ \#(\{i,j\} \cap \{k,l\}) = 1}} \sum_{t>l} C_n(\xi_{ij}, \xi_{tl})}_{A_2} + \underbrace{\sum_{\substack{j>i \\ \{i,j\} \cap \{k,l\} = \phi}} \sum_{t>l} C_n(\xi_{ij}, \xi_{tl})}_{A_3}\right\}.\end{aligned}$$

The number of terms in the summation A_1, A_2, A_3 are $\binom{n}{2} = \frac{1}{2}(n^2 - n)$, $\binom{n}{2} \times 2(n-2) = n^3 - 3n^2 + 2n$, and $\binom{n}{2} \times \binom{n-2}{2} = \frac{1}{4}(n^4 - 6n^3 + 11n^2 - 6n)$, respectively.

For A_1,

$$V_n(\xi_{ij}) = \frac{1}{4}[V_n(y_j \mathbb{1}'_j \iota_i g_{ij}) + V_n(y_i \mathbb{1}'_i \iota_j g_{ji}) + 2C_n(y_j \mathbb{1}'_j \iota_i g_{ij}, y_i \mathbb{1}'_i \iota_j g_{ji})].$$

Note that $E_n(y_j \mathbb{1}'_j \iota_i g_{ij}) \in [0,1]$ and $E_n(y_i \mathbb{1}'_i \iota_j g_{ji}) \in [0,1]$, then $V_n(\xi_{ij}) \in \frac{1}{4}\{[0,1] + [0,1] + 2[-1,1] \times [0,1]^2\} = [0,1]$. Thus, $\frac{4}{n^2(n-1)^2} A_1 \leq \frac{2}{n(n-1)}$.

For A_2 such that $\#(\{i,j\} \cap \{k,l\}) = 1$,

$$|C_n(\xi_{ij}, \xi_{tl})| = \left|\rho V_n^{1/2}(\xi_{ij}) V_n^{1/2}(\xi_{lt})\right| \leq 1$$

Thus, $|\frac{4}{n^2(n-1)^2} A_2| \leq \frac{4(n-2)}{n(n-1)}$.

For A_3 such that $\{i,j\} \cap \{k,l\} = \phi$,

$$\begin{aligned} C_n(\xi_{ij}, \xi_{tl}) =& \frac{1}{4} C_n\left(y_j \mathbb{1}'_j \iota_i g_{ij} + y_i \mathbb{1}'_i \iota_j g_{ji}, y_t \mathbb{1}'_t \iota_l g_{lt} + y_l \mathbb{1}'_l \iota_t g_{tl}\right) \\ =& \frac{1}{4}\{C_n\left(y_j \mathbb{1}'_j \iota_i g_{ij}, y_t \mathbb{1}'_t \iota_l g_{lt}\right) + C_n\left(y_j \mathbb{1}'_j \iota_i g_{ij}, y_l \mathbb{1}'_l \iota_t g_{tl}\right) \\ & + C_n\left(y_i \mathbb{1}'_i \iota_j g_{ji}, y_t \mathbb{1}'_t \iota_l g_{lt}\right) + C_n\left(y_i \mathbb{1}'_i \iota_j g_{ji}, y_l \mathbb{1}'_l \iota_t g_{tl}\right)\} \end{aligned}$$

Notice that

$$C_n\left(y_j \mathbb{1}'_j \iota_i g_{ij}, y_t \mathbb{1}'_t \iota_l g_{lt}\right) = E(y_j y_t \mathbb{1}'_j \mathbb{1}'_t \iota_i \iota_l g_{ij} g_{lt}) - E_n(y_j \mathbb{1}'_j \iota_i g_{ij}) E_n(y_t \mathbb{1}'_t \iota_l g_{lt})$$

and

$$
\begin{aligned}
&E(y_j y_t \mathbb{1}'_j \mathbb{1}'_t \iota_i \iota_l g_{ij} g_{lt} | x_j, \mathbf{d}_j, y_{N_j}, x_t, \mathbf{d}_t, y_{N_t}) \\
=&E[\mathbb{1}\{\varepsilon_j < h(x_j, \mathbf{d}_j) + \sum_{m \in N_i} \gamma(x_m, x_i, \mathbf{d}_i)(y_m + (1 - y_m)\sigma_{im})\} \\
&\cdot \mathbb{1}\{\varepsilon_t < h(x_t, \mathbf{d}_t) + \sum_{k \in N_t} \gamma(x_k, x_t, \mathbf{d}_t)(y_k + (1 - y_k)\sigma_{tk})\} \\
&\cdot \mathbb{1}\{x_j = x', x_t = x' x_i = x, x_l = x, \mathbf{d}_i = \mathbf{d}, \mathbf{d}_l = \mathbf{d}, g_{ij} = 1, g_{lt} = 1\} | x_j, \mathbf{d}_j, y_{N_j}, x_t, \mathbf{d}_t, y_{N_t}],
\end{aligned}
$$

Then $E(y_j y_t \mathbb{1}'_j \mathbb{1}'_t \iota_i \iota_l g_{ij} g_{lt} | x_j, \mathbf{d}_j, y_{N_j}, x_t, \mathbf{d}_t, y_{N_t}) = o(1/n)$ by Assumption 14. $\square$

2.11 C. Tables

Table 2.6: Estimating Payoff Parameters when ε is Uniformly Distributed

		Network Size					
		$n = 1000$			$n = 2000$		
	# networks	BIAS	SE	RMSE	BIAS	SE	RMSE
Uniform	$r = 1$	-0.0156	0.1292	0.1288	-0.0030	0.0898	0.0889
		-0.0404	0.1632	0.1666	-0.0169	0.0977	0.0982
		0.0638	0.3017	0.3054	0.0044	0.0495	0.0492
		0.0093	0.0689	0.0689	0.0012	0.0409	0.0405
		0.0228	0.1279	0.1287	0.0049	0.0714	0.0709
		0.1320	0.2722	0.3001	0.0246	0.1211	0.1224
	$r = 5$	0.0128	0.0501	0.0512	0.0038	0.0545	0.0541
		0.0016	0.0416	0.0412	0.0011	0.0296	0.0293
		0.0026	0.0347	0.0345	0.0069	0.0412	0.0414
		-0.0089	0.0222	0.0237	-0.0071	0.0188	0.0199
		0.0003	0.0176	0.0174	-0.0013	0.0154	0.0153
		0.0007	0.0310	0.0307	0.0000	0.0126	0.0125
	$r = 10$	0.0043	0.0545	0.0541	-0.0109	0.0480	0.0488
		-0.0009	0.0273	0.0270	0.0017	0.0189	0.0188
		0.0096	0.0430	0.0436	0.0132	0.0353	0.0373
		-0.0067	0.0163	0.0175	-0.0002	0.0194	0.0192
		0.0044	0.0142	0.0147	-0.0013	0.0105	0.0105
		0.0008	0.0203	0.0201	-0.0003	0.0078	0.0077
	$r = 20$	-0.0096	0.0439	0.0445	0.0008	0.0320	0.0317
		0.0031	0.0160	0.0161	-0.0026	0.0147	0.0148
		0.0175	0.0386	0.0420	0.0048	0.0293	0.0294
		-0.0019	0.0148	0.0147	-0.0034	0.0091	0.0097
		0.0043	0.0080	0.0090	0.0014	0.0077	0.0077
		-0.0027	0.0074	0.0078	0.0009	0.0062	0.0062

Note: The bias, standard deviation and root mean squared errors in this table are calculated from S = 50 independent draws of r networks, where each network has n individuals. $\varepsilon \sim U(0, 1)$.

Table 2.7: Estimating Payoff Parameters when ε is Normally Distributed

		Network Size					
		$n = 1000$			$n = 2000$		
	# networks	BIAS	SE	RMSE	BIAS	SE	RMSE
Normal	$r = 1$	0.0191	0.0854	0.0867	0.0113	0.0632	0.0636
		0.0114	0.0715	0.0717	0.0106	0.0440	0.0449
		−0.0076	0.0349	0.0354	−0.0054	0.0332	0.0333
		−0.0040	0.0442	0.0440	−0.0002	0.0299	0.0296
		0.0045	0.0423	0.0421	−0.0085	0.0218	0.0232
		−0.0057	0.0352	0.0353	−0.0035	0.0220	0.0220
	$r = 5$	0.0131	0.0377	0.0396	0.0034	0.0297	0.0296
		−0.0022	0.0305	0.0303	0.0034	0.0230	0.0230
		−0.0048	0.0191	0.0195	0.0002	0.0151	0.0149
		−0.0047	0.0208	0.0211	−0.0018	0.0138	0.0138
		0.0032	0.0159	0.0160	−0.0037	0.0127	0.0131
		−0.0005	0.0160	0.0158	−0.0009	0.0109	0.0108
	$r = 10$	0.0160	0.0354	0.0386	0.0009	0.0196	0.0194
		0.0029	0.0204	0.0204	0.0003	0.0173	0.0171
		−0.0046	0.0177	0.0181	−0.0010	0.0105	0.0104
		−0.0073	0.0155	0.0169	0.0001	0.0093	0.0092
		0.0007	0.0127	0.0126	−0.0006	0.0097	0.0096
		−0.0013	0.0108	0.0108	0.0013	0.0076	0.0076
	$r = 20$	0.0127	0.0183	0.0221	0.0043	0.0143	0.0148
		−0.0028	0.0122	0.0124	−0.0012	0.0114	0.0114
		−0.0031	0.0098	0.0102	−0.0026	0.0072	0.0076
		−0.0055	0.0085	0.0101	−0.0001	0.0071	0.0071
		0.0041	0.0077	0.0087	−0.0021	0.0074	0.0076
		0.0011	0.0066	0.0066	0.0012	0.0053	0.0054

Note: The bias, standard deviation and root mean squared errors in this table are calculated from S = 50 independent draws of r networks, where each network has n individuals. $\varepsilon \sim N(0.5, 0.3)$.

CHAPTER 3

The Application of Network Model in the Diffusion of Microfinance

3.1 Introduction

This chapter discusses how we can measure the network effect in a diffusion process, for example, product adoptions and the diffusion of rumors.

Although social effect has been a field of great interest for economists, there is little empirical evidence that fully take the impact of local network structure into account. Labor economists that study network effects commonly suppose that individual take-up decisions depend only on the take-up rate in the population or in their local neighborhood, not on specific links to their local neighbors. However, the impact of network can change with the specific local network structure faced by each agent in the network. To be more specific, a network connect an individual to neighbors of different type, and thus resulting in different social effects. Instead of treating all neighbors the same in the previous literature, we are going to estimate network effects that differs with the neighbor types.

We apply our model in Chapter 2 to data on participation of a microfinance program in India villages to describe the impact of neighbors on individual decisions. We need a structural model because the analysis of diffusion games depends crucially on the model that depicts information transmission. Our model provides crucial assumptions on the mechanism that drive the information to diffuse through social links. Also, it allows us to study the various network effect across different types of agents who cares about their neighbors' opinions. It depends on unknown equilibrium beliefs which specify agents expectations about their neighbors' decisions. Using participation data from 43 villages each includes about 200 villagers, we estimate these equilibrium beliefs and the network effects. The model estimates suggest that positive network effect exists among households in the villages, and we also find an evidence of homophily as households tend to value the decision of neighbors who have the same financial wellness more.

3.1.1 Related Works

Social interactions lead to correlated individual choices, and the correlation forms the basis of identification. With increasing interest in social determinants of individual behavior, the literature grows fast to address the arised identification problems. As a review of this literature, Durlauf and Ioannides (2010)[?] summarize the significant advances made in the identification of peer effects, neighborhood effects, or more generally, social effects. In addition, Bramoullé, Djebbari, and Fortin (2009)[?] and Blume, Brock, Durlauf, and Jayaraman (2015)[?] show that in the case that interactions are structured through a social network, the correlated behaviors also occur in games with either privately or commonly observable types.

There is a vast and still growing empirical literature that identifies the effects of agents' neighborhood on behavior and outcomes. For example, Katz and Lazarsfeld (1955) study how the opinions of the majority are shaped, they suggest a small subset of influential individuals play a big role in filtering and reinterpreting the mass media. Glaeser, Sacerdote and Scheinkman (1996) evaluate the role of social interactions in criminal behavior. Other influential works include Coleman (1966), Granovetter (1994), Foster and Rosenzweig (1995), Topa (2001) and Conley and Udry (2005).

There is not much empirical literature about the identification of preference in dynamical processes. Banerjee et al. (2013)[6] study the diffusion of microfinance in Indian villages using a simulated moment approach. In economic theory, Bass (1969)[8] provide a widely-used model where diffusion depends on the adoption rate in the population. Morris (2000)[44] starts with a finite and fixed network such that players' choice depends on their neighbors' choice in the previous period. Watts (2002)[61] studies large networks such that players make irreversible adoption decisions if the fraction of her neighbors who adopt exceed some threshold. The networks discussed in these two papers are homogeneous, where there is only a single type of nodes and edges. Sadler (2017)[?] models information diffusion in more complicated networks that consist of heterogeneous types of nodes and edges. The analysis reveals the positive local externalities from adoption decisions, indicating that standard simplified assumptions may lead to misguided predictions.

This chapter is also related to network games with private information on links and payoffs. It is empirically sound to assume incomplete information especially when the network involves a large population. As shown in Banerjee, Chandrasekhar, Duflo and Jackson

(2016)[?], community members in rural India have very limited information about the full network structure. This topic is also discussed by Goldsmith-Pinkham and Imbens (2013)[?], where their model assumes that many independent small games are observed, *i.e.,* the econometricians observe a large number of repeated games on a fixed network structure. In addition, Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2010)[23] studies homogeneous players with partial information on network structures. Eraslan and Tang (2017)[?] study the identification and estimation of large network games with private links and payoffs. They consider linear interactions and ignore the fact that for undirected networks players hold partial information about their neighbors' network, which greatly reduces the computational difficulties. Canen, Schwartz and Song (2017)[?] adopts a behavioral approach to model games on networks where agents only observe part of their neighbors' types.

3.2 The Data

We use data from Banerjee et al. (2013)[6]. The dataset contains survey information from 75 villages in rural India, among which 34 villages participated in the microfinance program. The survey questionnaire collects various data from both household and individual levels. For our analysis, we restrict attention to household-level data from the 34 villages in which the microfinance company operates.

We observe the household characteristics and the social network structure from the data. The household characteristics are drawn from the household survey, which contains questions about household wealth (e.g., the number of rooms in a household, number of beds, amenities). The network structure is generated from individual questionnaire, in which villages are asked about their personal relationships (who they visit, who they borrow money from, etc.). By aggregating the individual relationships to the household level, we generate an undirected, unweighted social network.

With a simple logistic regression on only household characteristics data, we get coefficients for covariates[1] and the constant. The results are shown in Table 3.1. This simple regression implies that number of rooms per capita is one key factor that affects the household participation decisions.

Table 3.1: Microfinance Participation on Household Characteristics

	Estimate	SE	pValue
(Intercept)	-1.2098	0.3218	0.0002
# Rooms	0.0070	0.0853	0.9348
# Beds	-0.2831	0.1426	0.0471
Electricity	0.1559	0.1228	0.2044
Latrine	0.1793	0.0804	0.0257
# Rooms per capita	-1.0232	0.3925	0.0091
# Beds per capita	1.1465	0.6557	0.0804

Note: The table summarizes the logistic regression results for a basic analysis of key factors that affects household participation decisions.

Table 3.2: Data Description

	All		Type 1	
	Mean	Std. Dev.	Mean	Std. Dev.
Number of Households	212.233	53.536	52.837	26.143
Degree	9.656	1.642	8.786	1.956
Betweenness Centrality	189.733	63.103	162.811	64.107
# Rooms	2.308	0.413	3.470	0.331
# Beds	0.878	0.455	1.271	0.598
Electricity	1.474	0.230	1.352	0.267
Latrine	2.429	0.293	2.164	0.380
# Rooms per capita	0.561	0.120	1.079	0.079
# Beds per capita	0.201	0.110	0.364	0.182
Take-Up Rate	0.194	0.082	0.126	0.080

Note: Sample includes 43 villages. Type-1 households are those with 70 percentile and upper number of rooms per capita

Table 3.2 provides descriptive statistics. Villages that participates in the microfinance program have an average of 212 households. We denote the households with 70 percentile and upper number of rooms per capita by *type-1 households*, the rest of the households by *type-0 households*. From the table, the degree and the betweenness centrality imply that type-1 households are not specifically different from other households in terms of network position.

[1]The covariates include number of rooms in a household, number of beds, whether the household has private/government/no electricity, whether the household has own/common/no latrine, number of rooms per capita, and number of beds per capita.

19.4% of households participate in microfinance, while the type-1 households take-up rate is only 12.6%, with a standard deviation of 8% across villages.

3.3 Model and Estimation

3.3.1 Model structure

We describe our diffusion model in this subsection, with emphasis on the adaptions we made to tailor it to the data. The model we adopt is the diffusion model with players who are allowed to make deferred decisions.

As researchers, we observe the participation decisions at the end of the diffusion $y_i \in \{0, 1\}$, the network structure G, and the types of households, type-0 and type-1, as divided in the previous section.

The diffusion is characterized by the following algorithm. At $t = 0$, a set of seed players (households) are informed about the new microfinance program. In each subsequent period, those who are informed decide whether to participate based on their own characteristics and the decisions of their close neighbors in the social network. When player participate, they transmit the information to their neighbors, so that their neighbors become aware of the new program. The decision to participate is irreversible, that is, players who already participated could not withdraw from the program.

Player i's payoff function is

$$U_i = h_i + \sum_{j \in N_i} \gamma_{ij} \cdot y_j - \varepsilon_i, \tag{3.1}$$

or

$$U_i = h_i + \sum_{j \in N_i} \gamma_{ij} \cdot y_j / d_i - \varepsilon_i, \tag{3.2}$$

where $h_i = h(x_i)$ depends on player i's type x_i and $\gamma = \gamma(x_j, x_i)$ depends on the types of both ends of the link x_i and x_j.

We assume players hold private information about their links and their payoff shocks ε.

Because of incomplete information, agents form beliefs about their neighbors' participation decisions σ. According to the theorems in the previous chapter, player i's belief about j's decision, σ_{ij}, depends on x_i, x_j and player i's degree $\mathbf{d}_i$. In this application, a player can have up tp about 50 neighbors in their social network. Therefore, to address the arising computational difficulty, we categorize the degree into three groups of the same sizes by the quantiles of the degree distribution (0.33 and 0.67).

3.3.2 Estimation and Bootstrap

We apply the two-step estimator estimate to estimate the payoff parameters. First, we estimate σ's from sample analogs. Then we take maximum-likelihood estimation. We use bootstrap to approach the variance of the estimate.

The bootstrap algorithm is as follows. Note that we have a sample of 43 villages $x_1, ..., x_{43}$ drawn from a distribution F from which we wish to estimate the payoff parameter θ using a statistic $\hat{\theta} = T(x_1, ..., x_n)$. Then we generate a large number of random samples from $x_1, ..., x_{43}$, compute θ from each sample, and compute the standard deviation of these estimates.

1. Set B as the number of bootstraps, V as the number of villages. Assign weight w_r^b to village r in b-th bootstrap, where $w_r^b = e_r^b / \bar{e}^b$, with e_r^b i.i.d. exp(1) random variables and $\bar{e}^b = \frac{1}{R} \sum_r e_r^b$.

2. Compute σ^b using the weighted sample averages of participating rate by types and degrees.

3. For each village r, compute the log-likelihood $l_r(\theta, \sigma^b)$ by plugging in σ^b computed in the first step.

4. Find $\theta^b = \arg\max L^b(\theta, \sigma^b)$, where $L^b(\theta, \sigma^b) = \frac{1}{R} \sum_r w_r^b \cdot l_r(\theta, \sigma^b)$.

5. Compute the standard deviation and confidence interval from $\{\theta^b\}_{b=1}^B$.

We use $B = 1000$ for bootstrap.

3.3.3 Comparison with the Mean-field Model

We compare our model to the canonical Bass model (Bass 1969[8]), which falls in the category of mean-field models. They do not consider local network structure because it relies on mean-field approximations, which assume all prior adopters in the population influence current decisions equally. The model characterize adoption rate over time using a logistic curve. In contrast, our model assumes fixed network with persistent relationships, where players react to their local neighbors' decisions strategically.

Figure 3.1 presents the different diffusion patterns between two models. We fitted the data to our model for the coefficients, and then perform simulations to generate the blue curve. Then we fit the simulated participation rate to the Bass model. [2] The two models generate significantly different diffusion patterns. The simulation suggests that the Bass model is not a good fit for the data as the local network plays an crucial role in the microfinance take-up.

3.4 Empirical Results

Table 3.3 presents the result of the estimation. The first row is the coefficient estimates of type-0 households' payoffs, and the second is the coefficient estimates of type-1 households' payoffs. The results demonstrate that social effect exists for households in the network, and are in general positive. The impact of type-0 neighbors on both type-0 and type-1 households are significantly different from zero, with $P < 0.01$ for type-0 on type-0, and $P < 0.05$ for type-0 on type-1 (t test). While the impact of type-1 neighbor on both type-0 and type-1 households are less significant ($P = 0.11$ for type-1 on type-0, and $P < 0.10$ for type-1 on type-1, t test).

We can find evidence of homophily from the results. The same-type network effects γ_{00} and γ_{11} has greater coefficient values than the cross-type network effects γ_{01} and γ_{10}.

Recall Table 3.2, the microfinance take-up rate of type-1 households is significantly below average. Note that type-1 households are those who have more rooms per capital. Their

[2] We take this approach because we do not have data on the participation rate over time.

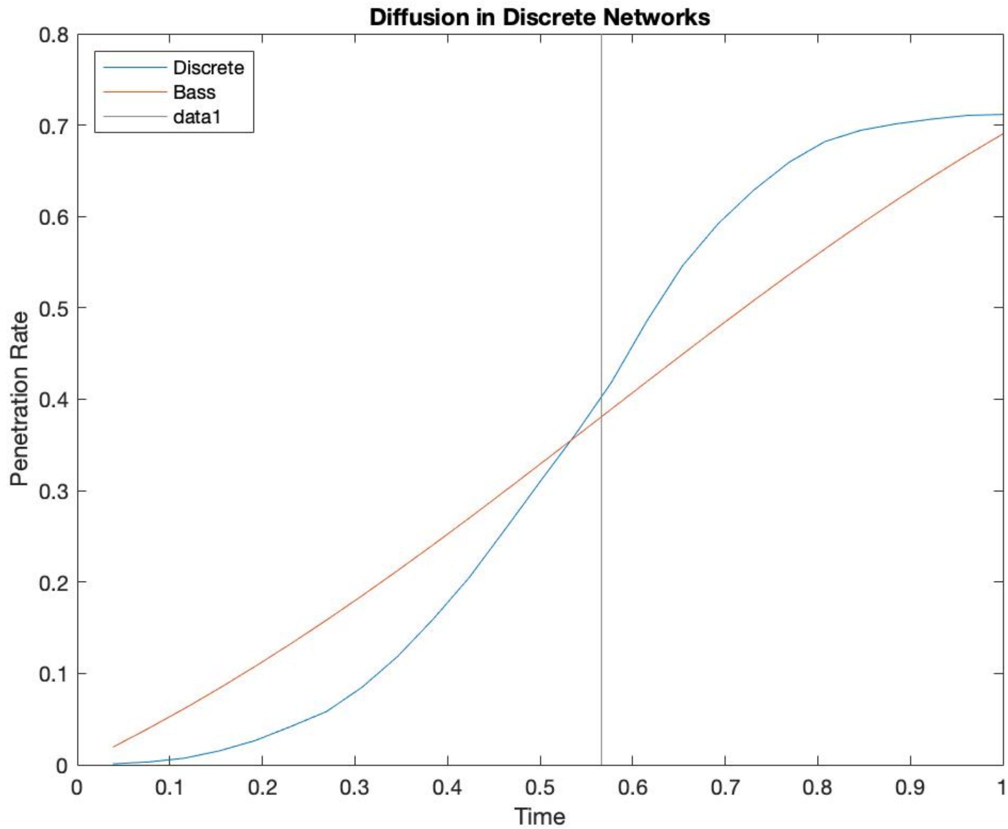


Figure 3.1: Fixed Network Approach v.s. Mean-field Model
Note: The figure demonstrates the diffusion pattern for two models. The blue curve presents the diffusion pattern when we fit the data to our model that based on the fixed network approach, and the brown curve is the diffusion pattern when we fit the data to a simple Bass model that makes mean-field assumptions.

advantage in wealth may therefore reduce their demand for financial instruments.

We also present the estimates of the nuisance parameter σ in Table 3.4. We observe two patterns from the table. First, the for each agent in the network, their beliefs about the participation decision of a type-1 neighbors is in general lower than that of a type-0 neighbor. This is consistent with the microfinance take-up rate in the data. Second, Households who have more links form lower beliefs for neighbor take-up rate, as the value of σ decreases along the diagonal.

Table 3.3: Parameter Estimates of the Model

	h_i	γ_{i0}	γ_{i1}
Type 0	0.3513	0.1200	0.0470
	(0.0091)	(0.0349)	(0.0396)
Type 1	0.3002	0.0619	0.1102
	(0.0114)	(0.0331)	(0.0806)

Note: This table presents the estimated payoff parameters. γ_{i0}, γ_{i1} are the coefficient in Equation (3.2) on the fraction of neighbors that participated. h_i is the constant term in Equation (3.2). Household of different types have different coefficients. Bootstrap estimation with 1000 draws is used to compute the standard errors of the parameter estimates.

3.5 Conclusion

This chapter applies the model in Chapter 2 to the data on the diffusion of microfinance from Banerjee et al. (2013)[6]. We find an evidence of positive network effect. To be more specific, the households with lower scores on financial wellness have more impact on their neighbors. Their neighbors benefit more from their decisions to participate in the microfinance program. We also find agents exhibit homophily.

Table 3.4: $\{\hat{\sigma}_{ij}\}$ by Household Types and Neighborhoods

		$j=0$			$j=1$		
	(d_0, d_1)	lower	middle	upper	lower	middle	upper
$i=0$	lower	0.3413	0.3708	0.3019	0.2274	0.3199	0.2602
		(0.0242)	(0.0416)	(0.0454)	(0.0432)	(0.0494)	(0.0504)
	middle	0.3468	0.3336	0.2884	0.1821	0.2560	0.2399
		(0.0304)	(0.0272)	(0.0263)	(0.0355)	(0.0326)	(0.0254)
	upper	0.3813	0.3007	0.2761	0.1493	0.2305	0.1877
		(0.0376)	(0.0299)	(0.0142)	(0.0463)	(0.0335)	(0.0183)
$i=1$	lower	0.3313	0.3476	0.2054	0.2118	0.2679	0.2867
		(0.0423)	(0.1007)	(0.0442)	(0.0728)	(0.0770)	(0.0626)
	middle	0.3566	0.2954	0.2926	0.3824	0.2620	0.1880
		(0.0362)	(0.0360)	(0.0324)	(0.0915)	(0.0541)	(0.0326)
	upper	0.3449	0.3225	0.2509	0.2686	0.2006	0.1972
		(0.0433)	(0.0402)	(0.0296)	(0.1193)	(0.0637)	(0.0280)

Note: This table presents agents' beliefs about the decisions of their neighbors, $\{\sigma_{ij}\}$, by household types (0,1) and neighborhood types (d_0, d_1 in the lower/middle/upper range, respectively).

REFERENCES

[1] Abrahamson, E., and Rosenkopf, L. Social network effects on the extent of innovation diffusion: A computer simulation. *Organization science 8*, 3 (1997), 289–309.

[2] Auerbach, E. Identification and estimation of a partially linear regression model using network data. *arXiv preprint arXiv:1903.09679* (2019).

[3] Bandiera, O., and Rasul, I. Social networks and technology adoption in northern mozambique. *The economic journal 116*, 514 (2006), 869–902.

[4] Banerjee, A., Breza, E., Chandrasekhar, A. G., Duflo, E., Jackson, M. O., and Kinnan, C. Changes in social network structure in response to exposure to formal credit markets. Tech. rep., National Bureau of Economic Research, 2021.

[5] Banerjee, A., Breza, E., Chandrasekhar, A. G., and Golub, B. When less is more: Experimental evidence on information delivery during india's demonetization. Tech. rep., National Bureau of Economic Research, 2018.

[6] Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. The diffusion of microfinance. *Science 341*, 6144 (2013).

[7] Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. Gossip: Identifying central individuals in a social network. Tech. rep., National Bureau of Economic Research, 2014.

[8] Bass, F. M. A new product growth for model consumer durables. *Management science 15*, 5 (1969), 215–227.

[9] Boswijk, H. P., and Franses, P. H. On the econometrics of the bass diffusion model. *Journal of Business & Economic Statistics 23*, 3 (2005), 255–268.

[10] Burlig, F., and Stephens, A. Reap what your friends sow: Social networks and technology adoption. Tech. rep., Mimeo, 2019.

[11] Cai, J., De Janvry, A., and Sadoulet, E. Social networks and the decision to insure. *American Economic Journal: Applied Economics 7*, 2 (2015), 81–108.

[12] Canen, N., Schwartz, J., and Song, K. Estimating local interactions among many agents who observe their neighbors. *Quantitative Economics 11*, 3 (2020), 917–956.

[13] Choo, E., and Siow, A. Who marries whom and why. *Journal of political Economy 114*, 1 (2006), 175–201.

[14] Coleman, J. S., Katz, E., and Menzel, H. Medical innovation: a diffusion study. inid% ianapolis, the bobbs% merril company. *Behavioral Science 12* (1966), 481–483.

[15] Conley, T. G., and Udry, C. R. Learning about a new technology: Pineapple in ghana. *American economic review 100*, 1 (2010), 35–69.

[16] Desiraju, R., Nair, H., and Chintagunta, P. Diffusion of new pharmaceutical drugs in developing and developed nations. *International Journal of Research in Marketing 21*, 4 (2004), 341–357.

[17] Diamond, W., and Agarwal, N. Latent indices in assortative matching models. *Quantitative Economics 8*, 3 (nov 2017), 685–728.

[18] Diamond, W., and Agarwal, N. Latent indices in assortative matching models. *Quantitative Economics 8*, 3 (2017), 685–728.

[19] Echenique, F., and Oviedo, J. A theory of stability in many-to-many matching markets.

[20] Eraslan, H., and Tang, X. Identification and estimation of large network games with private link information. Tech. rep., working paper, 2018.

[21] Foster, A. D., and Rosenzweig, M. R. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of political Economy 103*, 6 (1995), 1176–1209.

[22] Fox, J. T. Identification in matching games. *Quantitative Economics 1*, 2 (2010), 203–254.

[23] Galeotti, A., Goyal, S., Jackson, M. O., Vega-Redondo, F., and Yariv, L. Network games. *The review of economic studies 77*, 1 (2010), 218–244.

[24] Galichon, A., and Salanié, B. The econometrics and some properties of separable matching models. *American Economic Review 107*, 5 (2017), 251–55.

[25] Gatignon, H., Eliashberg, J., and Robertson, T. S. Modeling multinational diffusion patterns: An efficient methodology. *Marketing Science 8*, 3 (1989), 231–247.

[26] Glance, N. S., and Huberman, B. A. The outbreak of cooperation. *Journal of Mathematical sociology 17*, 4 (1993), 281–302.

[27] Goldenberg, J., Han, S., Lehmann, D. R., and Hong, J. W. The role of hubs in the adoption process. *Journal of marketing 73*, 2 (2009), 1–13.

[28] Griliches, Z. Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society* (1957), 501–522.

[29] Helsen, K., Jedidi, K., and DeSarbo, W. S. A new approach to country segmentation utilizing multinational diffusion patterns. *Journal of marketing 57*, 4 (1993), 60–71.

[30] Iyengar, R., Van den Bulte, C., and Valente, T. W. Opinion leadership and social contagion in new product diffusion. *Marketing Science 30*, 2 (2011), 195–212.

[31] Jackson, M. O., and López-Pintado, D. Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science 1*, 1 (2013), 49–67.

[32] Jackson, M. O., and Rogers, B. W. Relating network structure to diffusion properties through stochastic dominance. *BE Journal of Theoretical Economics 7*, 1 (2007), 1–16.

[33] Jackson, M. O., and Yariv, L. Diffusion on social networks. In *Economie Publique* (2005), Citeseer.

[34] Jackson, M. O., and Yariv, L. Diffusion of behavior and equilibrium properties in network games. *American Economic Review 97*, 2 (2007), 92–98.

[35] Jiao, Z., and Tian, G. The stability of many-to-many matching with max–min preferences. *Economics Letters 129* (2015), 52–56.

[36] Johnsson, I., and Moon, H. R. Estimation of peer effects in endogenous social networks: Control function approach. *Review of Economics and Statistics 103*, 2 (2021), 328–345.

[37] Konishi, H., and Ünver, M. U. Credible group stability in many-to-many matching problems. *Journal of Economic Theory 129*, 1 (2006), 57–80.

[38] Leskovec, J., Adamic, L. A., and Huberman, B. A. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB) 1*, 1 (2007), 5–es.

[39] Lohmann, S. The dynamics of informational cascades: The monday demonstrations in leipzig, east germany, 1989–91. *World politics 47*, 1 (1994), 42–101.

[40] López-Pintado, D. Contagion and coordination in random networks. *International Journal of Game Theory 34*, 3 (2006), 371–381.

[41] López-Pintado, D. Diffusion in complex social networks. *Games and Economic Behavior 62*, 2 (2008), 573–590.

[42] López-Pintado, D. Influence networks. *Games and Economic Behavior 75*, 2 (2012), 776–787.

[43] Menzel, K. Large matching markets as two-sided demand systems. *Econometrica 83*, 3 (2015), 897–941.

[44] Morris, S. Contagion. *The Review of Economic Studies 67*, 1 (2000), 57–78.

[45] Munshi, K. Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of development Economics 73*, 1 (2004), 185–213.

[46] Newman, M. Networks: An introduction.

[47] Peres, R., Muller, E., and Mahajan, V. Innovation diffusion and new product growth models: A critical review and research directions. *International journal of research in marketing 27*, 2 (2010), 91–106.

[48] Roth, A. E. Stability and polarization of interests in job matching. *Econometrica: Journal of the Econometric Society* (1984), 47–57.

[49] Roth, A. E. The origins, history, and design of the resident match. *Jama 289*, 7 (2003), 909–912.

[50] Ryan, B., and Gross, N. C. The diffusion of hybrid seed corn in two iowa communities. *Rural sociology 8*, 1 (1943), 15.

[51] Sadler, E. Diffusion games. *American Economic Review 110*, 1 (2020), 225–70.

[52] Schelling, T. C. Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. *Journal of Conflict resolution 17*, 3 (1973), 381–428.

[53] Schmittlein, D. C., and Mahajan, V. Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing science 1*, 1 (1982), 57–78.

[54] Skinner, J., and Staiger, D. *18. Technology Adoption from Hybrid Corn to Beta-Blockers*. University of Chicago Press, 2009.

[55] Sotomayor, M. Three remarks on the many-to-many stable matching problem. *Mathematical social sciences 38*, 1 (1999), 55–70.

[56] Srinivasan, V., and Mason, C. H. Nonlinear least squares estimation of new product diffusion models. *Marketing science 5*, 2 (1986), 169–178.

[57] Sundararajan, A. Local network effects and complex network structure. *Available at SSRN 650501* (2006).

[58] Takada, H., and Jain, D. Cross-national analysis of diffusion of consumer durable goods in pacific rim countries. *Journal of Marketing 55*, 2 (1991), 48–54.

[59] Talukdar, D., Sudhir, K., and Ainslie, A. Investigating new product diffusion across products and countries. *Marketing Science 21*, 1 (2002), 97–114.

[60] Van der Lans, R., Van Bruggen, G., Eliashberg, J., and Wierenga, B. A viral branching model for predicting the spread of electronic word of mouth. *Marketing science 29*, 2 (2010), 348–365.

[61] Watts, D. J. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences 99*, 9 (2002), 5766–5771.