Estimation of Large Network Formation Games*

Geert Ridder[†] Shuyang Sheng[‡]

December 5, 2016

Abstract

This paper develops estimation methods for network formation using observed data from a single large network. We characterize network formation as a simultaneous-move game with incomplete information, where we allow for utility externalities from indirect friends such as friends in common, so the expected utility from direct friends can be nonlinear. Nonlinearity poses a challenge in estimation because each individual faces an interdependent multinomial discrete choice problem with 2^{n-1} alternatives, which is difficult to solve. We propose a novel method to linearize the expected utility using Legendre transform and derive a closed-form expression for the conditional choice probability (CCP). Using the closed-form expression, we show that the CCP in n-player games converge to the CCP in a limit game as n approaches infinity. We propose a two-step estimation procedure that requires few assumptions on equilibrium selection, is simple to compute, and provides asymptotically valid estimators for the parameters. Monte Carlo results show that the estimation procedure can provide accurate estimates if networks are large.

KEYWORDS: Network formation, Large games, Incomplete information, Two-step estimation, Legendre Transform

JEL Codes: C13, C31, C57, D85

*We are grateful to Denis Chetverikov, Aureo de Paula, Bryan Graham, Jin Hahn, Bo Honore, Zhipeng Liao, Rosa Matzkin, Kevin Song and participants of seminars at UC Berkeley, UC Davis, UCLA, UC Riverside, Princeton, U Colorado, Northwestern, Econometric Society World Congress (2015), Econometric Society Winter Meeting and Summer Meeting (2016), INET conferences on econometrics of networks at University of Cambridge (2015) and USC (2015), Berkeley/CeMMAP conference on networks at Berkeley (2016).

†Department of Economics, University of Southern California, Los Angeles, CA 90089. E-mail: ridder@usc.edu.

‡Department of Economics, University of California at Los Angeles, Los Angeles, CA 90095. E-mail: ssheng@econ.ucla.edu.

1 Introduction

This paper contributes to the as yet small literature on the estimation of game-theoretic models of network formation.[1] The purpose of the empirical analysis is to recover the preferences of the members of the network, in particular the preferences that determine whether a member of the network will form a link (friendship, business relation or some other type of link) with a specific other member of the network. The preference for a link depends in general on the characteristics of the two members, and on their position in the network, e.g., their number of friends, their number of common friends. It is the dependence of the link preference on the position in the network that complicates the analysis. The preference also depends on unobservable features of the members and the link and assumptions on the nature of these unobservables play a key role in the empirical analysis.

In principle link formation models are discrete choice models where multiple alternatives (links) can be chosen. If the agent has a myopic strategy s/he chooses to form a link if the utility of the link is larger than the utility of not forming the link. There are two reasons why such a strategy is suboptimal. First, the agent is better off considering the choice of links with the other members in the network simultaneously. Second all members in the network are making link choices and all these choices have to be consistent. In this paper the consistency is achieved by assuming that the observed network is a Bayesian Nash equilibrium.

In general there is no unique Bayesian Nash equilibrium. This implies that full-information methods, either have to specify an equilibrium selection mechanism or have to consider partial instead of point identification of the preference parameters. In this paper we propose a limited-information method that is valid even if we do not know the equilibrium selection mechanism.

Assumptions regarding the unobservables in the link preferences play an important role in the empirical analysis. The extreme assumptions are complete information where all members (but not the econometrician) know the unobservables in the preferences for links with all members and incomplete information where a member only knows its own link specific unobservables. The complete information models are the hardest to estimate and they achieve set and not point identification of the parameters (Miyauchi (2013) and Sheng (2016)). Leung (2015) considers a model in which an agent only knows the link unobservable when considering to form that link. In this paper we consider a case in which the agent knows her/his own unobserved link preferences, but not those of other agents. So our assumption is between that in Leung (2015) and complete information, and our assumption is also in line with the usual assumption in discrete choice models.

[1] Jackson (2008) surveys game-theoretic models of network formation.

A further distinction in the empirical literature regards the data. We can have data on a large number of small networks, e.g., friendship networks in classrooms, or we can have data on a single large network. This paper considers the latter case (Menzel (2016b), Leung (2015), and De Paula, Richards-Shubik and Tamer (2015) also consider large networks). An advantage of the large single network case is that for a fairly general utility function under a Bayesian Nash equilibrium the link choices converge to the myopic decision rule of a single agent, because the normalized preferences converge to the preferences for this case. This is true even though the link preferences depend on a non-trivial and non-vanishing way on the position of the agent in the network. This simplification only holds if we add a 'sufficient statistic' that captures network position. An implication of this result is that we can estimate preference parameters by a two-step procedure. Since this procedure only uses the 'first-order condition' for optimal link choice it does not require an assumption on equilibrium selection beyond the assumption that the equilibrium selection does not change with the sample size, so that the first stage converges.

The plan of the paper is as follows. In Section 2 we introduce the model and the specific utility function that we will use. We also discuss the Bayesian Nash equilibrium for the network. In Section 3 we obtain a closed-form expression for the link-formation probability that avoids the solution of an integer program. We discuss in Section 4 the (uniform) convergence of the choice probabilities if the network size grows without bounds. Section 5 discusses the two-step estimator. Section 6 considers the extension to undirected networks and Section 7 conducts a simulation study.

2 Model

Let $\mathcal{N}_n = \{1, \ldots, n\}$ be a set of individuals who play to form links. The links form a network, which we denote by $G \in \mathcal{G}$. This is an $n \times n$ binary matrix. The (i, j) element $G_{ij} = 1$ if i,j are linked and 0 otherwise. The diagonal elements G_{ii} are set to be 0. We first consider directed networks, i.e., G_{ij} and G_{ji} may be different. The case of undirected links is discussed later in Section 6.

Each individual i has a vector of observed characteristics $X_i \in \mathcal{X}$ and a vector of unobserved preference shocks $\varepsilon_i = (\varepsilon_{ij})_{j \in \mathcal{N}_n} \in \mathbb{R}^n$, where ε_{ij} is i's preference for link ij and $\varepsilon_{ii} = 0$. We assume that the characteristics $X = (X_i)_{i \in \mathcal{N}_n}$ are publicly observed by all the individuals, but the shocks ε_i are private information of i. We also assume that the private shocks are i.i.d. and independent of the observables.

Assumption 1 *(i)* $\{\varepsilon_{ij}, i \neq j \in \mathcal{N}_n\}$, *are i.i.d. with CDF* $F(\theta_\varepsilon)$ *supported on* $\mathbb{R}^{n-1}$ *that is* *absolutely continuous with respect to Lebesgue measure. $F(\theta_\varepsilon)$ is known up to $\theta_\varepsilon \in \Theta_\varepsilon \subset \mathbb{R}^{d_\varepsilon}$. (ii) $\{X_i, i \in \mathcal{N}_n\}$ are i.i.d. (iii) For all $i \in \mathcal{N}_n$, ε_i and X are independent.*

Utility Given the network G, characteristic profile X, and private shocks ε_i, individual i has utility

$$\begin{aligned} U_i(G, X, \varepsilon_i; \theta) \quad = \quad & \frac{1}{n-1} \sum_{j \neq i} G_{ij} \left(v_{ij}(G_{-i}, X; \beta) - \varepsilon_{ij} \right) \\ & + \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} G_{ij} G_{ik} h_i \left(G_j, G_k; \gamma \right) \end{aligned} \tag{1}$$

where $G_i = (G_{ij})_{j \in \mathcal{N}_n}$ is the ith row of G (i.e., the links formed by i) and $G_{-i} = (G_j)_{j \neq i}$ is the submatrix of G with the ith row deleted (i.e., the links formed by individuals other than i). We assume that the utility is known up to $\theta = (\beta, \gamma)$ in a compact set $\Theta_u \subset \mathbb{R}^{d_u}$.

In the utility specification, we use $v_{ij}(G_{-i}, X; \beta)$ to represent the incremental utility from a link that is separable in i's links. A typical example of $v_{ij}(G_{-i}, X; \beta)$ could be

$$v_{ij}(G_{-i}, X; \beta) = \beta_0 + X_i' \beta_1 + |X_i - X_j|' \beta_2 + G_{ji} \beta_3 + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk} w(X_i, X_k) \beta_4 \tag{2}$$

where the first four terms capture the direct utility from j, while the last term captures the indirect utility from j's friends, in terms of j's weighted average out-degree, with $w(X_i, X_k)$ being the weight. The specification of separable indirect utility is similar to that in Leung (2015). What differentiates our work from his is that in addition to the separable indirect utility, we also allow for utility from indirect friends that are nonseparable in one's links. Such nonseparable indirect utility is represented by $h_i(G_j, G_k; \gamma)$. In fact, many well-known network externalities are nonseparable in one's links. For example, if we want to capture the utility from friends in common, we may specify

$$h_i(G_j, G_k; \gamma) = G_{jk} G_{kj} \gamma_1 + \frac{1}{n-3} \sum_{l \neq i,j,k} G_{jl} G_{kl} \gamma_2 \tag{3}$$

where the first term captures the effect of friends in common that are directly connected and the second term captures the effect of friends in common that are indirectly connected. It is essential to be able to allow for such nonseparable externalities if we want to explain clustering, a well-documented feature of social and economic networks (Jackson, 2008).

We normalize the summations appeared above by n or n^2 to ensure that those terms remain bounded when the number of individuals increases to infinity, the asymptotic scenario we will consider for estimation.

Equilibrium Let $G_i(X,\varepsilon_i)$ denote individual i's (pure) strategy, which maps i's information (X,ε_i) to a vector of links in $\mathcal{G}_i = \{0,1\}^{n-1}$. The optimal strategy of individual i maximizes her expected utility $\mathbb{E}\left[U_i(g_i,G_{-i},X,\varepsilon_i)\middle| X,\varepsilon_i\right]$ over $g_i \in \mathcal{G}_i$, where the expectation is taken with respect to others' actions G_{-i} given her belief about G_{-i}. Because the private shocks are independent by Assumption 1, individual i's belief about G_{-i} depends only on the public information X. Let $\sigma_i(g_i|X) = \Pr(G_i(X,\varepsilon_i) = g_i|X)$ be the probability that individual i chooses g_i given X. Under Assumption 1, actions $\{G_i, i\in\mathcal{N}_n\}$ are conditional independent given X, so individual i's belief about others' actions is $\sigma_{-i}(g_{-i}|X) = \prod_{j\neq i}\sigma_j(g_j|X)$. Let $\sigma(X) = \{\sigma_i(g_i|X), i\in\mathcal{N}_n, g_i\in\mathcal{G}_i\}$ be the belief profile, a function that maps characteristics $X\in\mathcal{X}^n$ to a $n2^{n-1}$-dimensional vector with entry values in $[0,1]$. Given belief profile σ, individual i's expected utility is

$$
\begin{aligned}
\mathbb{E}\left[U_i(G_i,G_{-i},X,\varepsilon_i)\middle| X,\varepsilon_i,\sigma\right] &= \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(\mathbb{E}\left[v_{ij}(G_{-i},X)\middle| X,\sigma\right]-\varepsilon_{ij}\right) \\
&\quad + \frac{1}{(n-1)(n-2)}\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}G_{ik}\mathbb{E}\left[h_i(G_j,G_k)\middle| X,\sigma\right] \\
&= \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(\sum_{g_{-i}} v_{ij}(g_{-i},X)\prod_{j\neq i}\sigma_j(g_j|X)-\varepsilon_{ij}\right) \\
&\quad + \frac{1}{(n-1)(n-2)}\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}G_{ik}\left(\sum_{g_j,g_k} h_i(g_j,g_k)\sigma_j(g_j|X)\sigma_k(g_k|X)\right)
\end{aligned}
\tag{4}
$$

and thus the probability that i chooses g_i given X and σ is

$$
\begin{aligned}
& \Pr(G_i = g_i|X,\sigma) \\
= \; & \Pr\left(\mathbb{E}\left[U_i(g_i,G_{-i},X,\varepsilon_i)\middle| X,\varepsilon_i,\sigma\right] \geq \max_{\tilde{g}_i\in\mathcal{G}_i}\mathbb{E}\left[U_i(\tilde{g}_i,G_{-i},X,\varepsilon_i)\middle| X,\varepsilon_i,\sigma\right]\middle| X,\sigma\right).
\end{aligned}
\tag{5}
$$

In a n-player game, a Bayesian Nash equilibrium $\sigma^*(X)$ is a belief profile that satisfies

$$
\sigma_i^*(g_i|X) = \Pr(G_i = g_i|X,\sigma^*(X)) \tag{6}
$$

for all characteristic profiles $X\in\mathcal{X}^n$, actions $g_i\in\mathcal{G}_i$, and individuals $i\in\mathcal{N}_n$.

In this paper, we focus on equilibria that are symmetric in individuals' observed characteristics. We say an equilibrium $\sigma(X)$ is *symmetric* if for i and j with $X_i = X_j$, we have $\sigma_i(X) = \sigma_j(X)$. In network data where individuals do not have identities, it makes sense to assume symmetric equilibria because otherwise the optimal strategies and equilibria may depend on how we label the individuals. More importantly, as we assume only one network is observed, the assumption of symmetric equilibria will be crucial for valid estimation and inference. It can be shown that there exists a symmetric equilibrium. We assume that the network observed is a symmetric equilibrium.

Proposition 1 *For any $X \in \mathcal{X}^n$, there exists a symmetric equilibrium $\sigma(X)$.*

Proof. See the appendix. ■

3 Representation of the Optimal Strategies

Analyzing the choice probability in (5), especially its asymptotic behavior when the number of individuals n approaches infinity, is challenging because the expected utility in (4) is nonseparable in G_i, while the entries of G_i take binary values, so there is no simple solution for the optimal G_i. In this section, we propose a novel approach to deriving the optimal G_i in a pseudo closed form. This pseudo closed form is simple enough that it enables us to explicitly characterize the dependence between the entries in the optimal G_i, derive the limit of the choice probability in (5) as n approaches infinity, and propose estimates for the choice probabilities that are asymptotically valid based on a single observed network. These properties are crucial for the estimation for model parameters.

The insight of our approach is to transform the expected utility into a form with auxiliary variables that is linear in G_i, so the optimal G_i can be solved in closed form. To facilitate the presentation of the idea, we assume that X_i has a finite support so the derivation can be presented in simple matrix notation. This assumption is inessential and can be relaxed to allow for continuous X_i.

Assumption 2 *The support of X_i has finite distinct values, $\mathcal{X} = \left\{x^1, \ldots, x^T\right\}$.*

To proceed, first observe that by symmetric assumption the quadratic form in the expected utility in (5) is symmetric in j and k. Moreover, in a symmetric equilibrium $\sigma(X)$, if two individuals j and k have the same characteristics, i.e., $X_j = X_k$, we have $\sigma_j(X) = \sigma_k(X)$. This implies that the (j,k) element in the matrix obtained from the quadratic form, i.e., $\mathbb{E}\left[h_i(G_j, G_k)\middle|\,\sigma\right] = \sum_{g_j,g_k} h_i(g_j, g_k)\,\sigma_j(g_j|\,X)\,\sigma_k(g_k|\,X)$, depends on (j,k) only through the value of (X_j, X_k). In other words, for two pairs of individuals (j,k) and (l,m) with the same characteristics $(X_j, X_k) = (X_l, X_m)$, we have $\mathbb{E}\left[h_i(G_j, G_k)\middle|\,\sigma\right] = \mathbb{E}\left[h_i(G_l, G_m)\middle|\,\sigma\right]$. Therefore, given $X_{-jk} = (X_l)_{l \neq j,k}$, we can view $\mathbb{E}\left[h_i(G_j, G_k)\middle|\,\sigma\right]$ as a function of (X_j, X_k), and denote the value of this function at $(X_j, X_k) = (x^s, x^t)$, $1 \leq s, t \leq T$, by

$$\begin{aligned} H_{st}(\sigma(X)) &= \mathbb{E}\left[h_i(G_j, G_k) \middle| \sigma\left(X_j = x^s, X_k = x^t, X_{-jk}\right)\right] \\ &= \sum_{g_j, g_k} h_i(g_j, g_k)\, \sigma_j\left(g_j \middle| X_j = x^s, X_k = x^t, X_{-jk}\right) \sigma_k\left(g_k \middle| X_j = x^s, X_k = x^t, X_{-jk}\right) \end{aligned}$$

Write all such values in a matrix

$$H(\sigma(X)) = \begin{bmatrix} H_{11}(\sigma(X)) & \cdots & H_{1T}(\sigma(X)) \\ \vdots & & \vdots \\ H_{T1}(\sigma(X)) & \cdots & H_{TT}(\sigma(X)) \end{bmatrix}$$

The above defined $H_{st}(\sigma)$ and $H(\sigma)$ may have subscript i but we abbreviate it for simplicity because whenever we consider them we will fix individual i.

Example 1 *For the nonseparable utility specification in (3), we have*

$$\mathbb{E}\left[h_i(G_j, G_k) \middle| X, \sigma\right] = \sigma_{jk}(X)\, \sigma_{kj}(X)\, \gamma_1 + \frac{1}{n-3} \sum_{l \neq i,j,k} \sigma_{jl}(X)\, \sigma_{kl}(X)\, \gamma_2$$

where $\sigma_{jk}(X) = \Pr\left(G_{jk} = 1 \middle| X\right)$, $j, k \in \mathcal{N}_n$, *so*

$$\begin{aligned} H_{st}(\sigma(X_{-jk})) &= \sigma_{jk}\left(x^s, x^t, X_{-jk}\right) \sigma_{kj}\left(x^s, x^t, X_{-jk}\right) \gamma_1 \\ &\quad + \frac{1}{n-3} \sum_{l \neq i,j,k} \sigma_{jl}\left(x^s, x^t, X_{-jk}\right) \sigma_{kl}\left(x^s, x^t, X_{-jk}\right) \gamma_2 \end{aligned}$$

As discussed $H(\sigma)$ is a real symmetric matrix. It has a real spectral decomposition

$$H(\sigma) = \Phi(\sigma)\, \Lambda(\sigma)\, \Phi(\sigma)' \tag{7}$$

where $\Lambda(\sigma) = diag(\lambda_1(\sigma), \ldots, \lambda_T(\sigma))$ is the diagonal matrix of eigenvalues and $\Phi(\sigma) = (\phi_1(\sigma), \ldots, \phi_T(\sigma))$ is the orthogonal matrix of eigenvectors. Using the eigendecomposition in (7), we can reduce the quadratic form in (4) to a canonical quadratic form that involves only squares of linear functions of G_i. Once we have the canonical form, we can linearize those squares of linear functions of G_i by applying a special case of Legendre transform (Rockafellar, 1970) for square functions, namely,

$$Y^2 = \max_{\omega \in \mathbb{R}} \left\{ 2Y\omega - \omega^2 \right\} \tag{8}$$

Note that the maximand on the right hand side of (8) is linear in Y. The geometric intuition of (8) is that a square function can be approximated by the maximum of its support functions which are linear. Replacing those squares by their counterparts as in the right hand side of (8), we derive the expected utility in a form that are linear in G_i.[2] The results are summarized in Proposition 2.

Proposition 2 *Under Assumptions 1-2, the expected utility in (4) is equal to*

$$\begin{aligned} \mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] &= \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(V_{n,ij}\left(X,\sigma\right) - \varepsilon_{ij}\right) \\ &\quad + \frac{n-1}{n-2}\sum_{t}\lambda_t\left(\sigma\right)\left(\frac{1}{n-1}\sum_{j\neq i} G_{ij} D_j' \phi_t\left(\sigma\right)\right)^2 \end{aligned}$$

with

$$V_{n,ij}\left(X,\sigma\right) = \sum_{g_{-i}} v_{ij}(g_{-i}, X)\prod_{j\neq i}\sigma_j\left(g_j\middle| X\right) - \frac{1}{n-2}D_j' diag\left(H\left(\sigma\right)\right)D_j \tag{9}$$

$$D_i = \left(1\left\{X_i = x^1\right\}, \ldots, 1\left\{X_i = x^T\right\}\right)'$$

so by Legendre transform in (8)

$$\begin{aligned} \mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] &= \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(V_{n,ij}\left(X,\sigma\right) - \varepsilon_{ij}\right) \\ &\quad + \frac{n-1}{n-2}\sum_{t}\lambda_t\left(\sigma\right)\max_{\tilde{\omega}_t\in\mathbb{R}}\left\{2\left(\frac{1}{n-1}\sum_{j\neq i} G_{ij} D_j'\phi_t\left(\sigma\right)\right)\tilde{\omega}_t - \tilde{\omega}_t^2\right\} \end{aligned} \tag{10}$$

Proof. See the appendix. ■

The optimal G_i is solved by maximizing the expected utility. Given the equivalent representation of the expected utility in (10), if we can interchange the inner maximization over $\tilde{\omega} = (\tilde{\omega}_1, \ldots, \tilde{\omega}_T) \in \mathbb{R}^T$ with the outer maximization over G_i, then because the maximand is linear in G_i, the optimal G_i can be solved easily. To achieve this, we first need to move the inner maximization operation outside the entire expected utility function. This is ensured by the assumption that the eigenvalues of $H\left(\sigma\right)$ are nonnegative. Lemma 1 then guarantees the validity of the swap of the two maximization operations and the uniqueness of the solution in the swapped problem if the solution to the original problem is unique. Moreover, define $\omega = (\omega_1, \ldots, \omega_T) = \Phi\tilde{\omega} \in \mathbb{R}^T$. By (7) and $\Phi^{-1} = \Phi'$, we have $H\left(\sigma\right)\omega = \Phi\left(\sigma\right)\Lambda\left(\sigma\right)\tilde{\omega}$ and $\omega' H\left(\sigma\right)\omega = \tilde{\omega}'\Lambda\left(\sigma\right)\tilde{\omega}$, so the maximization over $\tilde{\omega}$ can be replaced by an equivalent maximization over ω. It is easier to work with matrix $H\left(\sigma\right)$ than its eigenvalues and eigenvectors

[2] We are grateful to Terrence Tao for suggesting the idea of Legendre transform.

because the latter may not be unique.

Assumption 3 (Positive externality from clustering) *Any symmetric equilibrium* σ_n^* *of (6) has a neighborhood such that any symmetric* σ *in the neighborhood satisfies* $\min_t \lambda_t(H(\sigma)) \geq 0$.

Lemma 1 *For any function* $f(x,y): \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ *with* $\sup_{x,y} f(x,y) < \infty$, *we have*

$$\max_y \max_x f(x,y) = \max_x \max_y f(x,y) \tag{11}$$

Therefore, if there is a unique (x^*, y^*) *such that* $f(x^*, y^*) = \max_y \max_x f(x,y)$, *then* (x^*, y^*) *is also the unique solution to* $\max_x \max_y f(x,y)$.

The equivalent transformations are summarized below

$$\begin{aligned}
&\max_{G_i} && \mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\
\Leftrightarrow &\max_{G_i}\max_{\tilde{\omega}} && \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'\sum_t \phi_t(\sigma)\lambda_t(\sigma)\tilde{\omega}_t - \varepsilon_{ij}\right) - \frac{n-1}{n-2}\sum_t \lambda_t(\sigma)\tilde{\omega}_t^2 \\
\Leftrightarrow &\max_{\tilde{\omega}}\max_{G_i} && \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'\Phi(\sigma)\Lambda(\sigma)\tilde{\omega} - \varepsilon_{ij}\right) - \frac{n-1}{n-2}\tilde{\omega}'\Lambda(\sigma)\tilde{\omega} \\
\Leftrightarrow &\max_{\omega}\max_{G_i} && \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'H(\sigma)\omega - \varepsilon_{ij}\right) - \frac{n-1}{n-2}\omega'H(\sigma)\omega
\end{aligned} \tag{12}$$

It is immediate to derive the optimal G_i from (12).

Theorem 1 *Under Assumptions 1-3, for any* (X, ε_i) *and symmetric belief profile* σ *in a neighborhood of a symmetric equilibrium* σ_n^*, *the optimal strategy* $G_{n,i}(X,\varepsilon_i,\sigma) = (G_{n,ij}(X,\varepsilon_i,\sigma))_{j\neq i}$ *is given by*

$$G_{n,ij}(X,\varepsilon_i,\sigma) = 1\left\{V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'H(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma) - \varepsilon_{ij} \geq 0\right\},\ \forall j \neq i \tag{13}$$

where $\omega_{n,i}(X,\varepsilon_i,\sigma)$ *is a maximizer of the problem*

$$\max_{\omega} \Pi_n(\omega, X, \varepsilon_i, \sigma) = \frac{1}{n-1}\sum_{j\neq i}\left[V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+ - \frac{n-1}{n-2}\omega'H(\sigma)\omega \tag{14}$$

with the notation $[x]_+ = \max\{x, 0\}$. *Moreover, the optimal* $G_{n,ij}(X,\varepsilon_i,\sigma)$ *and* $H(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)$ *are unique almost surely.*

Proof. It suffices to show the uniqueness statement. Under the assumption that ε_i has a continuous distribution, the optimal $G_{n,i}(X, \varepsilon_i, \sigma)$ solved from the original problem in (5) is unique almost surely, so by Lemma 1 the optimal $G_{n,i}(X, \varepsilon_i, \sigma)$ given in (13) is unique almost surely. The almost sure uniqueness of $H(\sigma)\,\omega_{n,i}(X, \varepsilon_i, \sigma)$ is immediate. ■

4 Convergence of the Game

In this section, we explore the asymptotic behavior of the n-player game when n approaches infinity. Thanks to the closed form representation in (13)-(14), we can show that the optimal strategy of an individual and the probability she forms a link converge to some limiting strategy and link formation probability as n goes to infinity. The optimal strategy in the limit has a simple form that conditional on some sufficient statistics that control for the spillover effects between one's own links, the links of an individual are formed independently. These asymptotic features are crucial for deriving simple enough while asymptotically valid estimators of the model parameters.

From (13) the probability that individual i forms a link to j conditional on characteristic profile X and belief profile σ is

$$\begin{aligned} P_n(X_i, X_j; X, \sigma) &= \Pr\left(G_{n,ij}(X, \varepsilon_i, \sigma) = 1 \middle| X, \sigma\right) \\ &= \Pr\left(V_{n,ij}(X, \sigma) + \frac{n-1}{n-2} 2D_j' H(\sigma)\,\omega_{n,i}(X, \varepsilon_i, \sigma) - \varepsilon_{ij} \geq 0 \middle| X, \sigma\right) \end{aligned} \tag{15}$$

where the probability operation is calculated with respect to ε_i. We write (X_i, X_j) explicitly in $P_n(X_i, X_j; X, \sigma)$ to emphasize that by symmetry this link formation probability depends on (i, j) only through characteristics (X_i, X_j). If the utility specification ensures that $V_{n,ij}(X, \sigma)$ converges to some limit $V_{ij}(\sigma)$ that depends on (X_i, X_j) as $n \to \infty$, as stated in Assumption 4, we expect that the link formation probability $P_n(X_i, X_j; X, \sigma)$ converges to a limit given by

$$P(X_i, X_j; \sigma) = \Pr\left(V_{ij}(\sigma) + 2D_j' H(\sigma)\,\omega_i(\sigma) - \varepsilon_{ij} \geq 0 \middle| X_i, X_j, \sigma\right) \tag{16}$$

where $\omega_i(\sigma)$ is a maximizer of the problem

$$\max_{\omega} \Pi(\omega, X_i, \sigma) = \mathbb{E}\left(\left[V_{ij}(\sigma) + 2D_j' H(\sigma)\,\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right) - \omega' H(\sigma)\,\omega \tag{17}$$

as $n \to \infty$. The expectation in (17) is taken with respect to X_j and ε_{ij}. The equilibrium belief profile σ^* is a fixed point solved from

$$\sigma_{ij}^*(X_i, X_j) = P(X_i, X_j; \sigma^*)$$

where $\sigma_{ij}(X_i, X_j) = \Pr(G_{ij} = 1 | X_i, X_j)$.

Assumption 4 *For $V_{n,ij}(X, \sigma)$ defined in (9), there is $V_{ij}(\sigma) = V(X_i, X_j, \sigma)$ such that*

(a) given any (X_i, X_j), $\sup_\sigma |V_{n,ij}(X, \sigma) - V_{ij}(\sigma)| \xrightarrow{p} 0$, and

(b) given any X_i, $\sup_\sigma \frac{1}{n-1} \sum_{j \neq i} |V_{n,ij}(X, \sigma) - V_{ij}(\sigma)| \xrightarrow{p} 0$.

Example 2 *Consider the separable utility specification in (2). For any X and symmetric σ, we have*

$$\begin{aligned} V_{n,ij}(X, \sigma) &= \beta_0 + X_i'\beta_1 + |X_i - X_j|'\beta_2 + \sigma_{ji}(X_j, X_i)\beta_3 \\ &\quad + \frac{1}{n-2} \sum_{k \neq i,j} \sigma_{jk}(X_j, X_k) w(X_i, X_k)\beta_4 - \frac{1}{n-2} D_j' diag(H(\sigma)) D_j \end{aligned}$$

We verify in the appendix that $V_{ij}(\sigma)$ given by

$$\begin{aligned} V_{ij}(\sigma) &= \beta_0 + X_i'\beta_1 + |X_i - X_j|'\beta_2 + \sigma_{ji}(X_j, X_i)\beta_3 \\ &\quad + \mathbb{E}[\sigma_{jk}(X_j, X_k) w(X_i, X_k) | X_i, X_j] \beta_4 \end{aligned}$$

satisfies Assumption 4.

We refer to $P(X_i, X_j; \sigma)$ defined in (16) the limiting choice probability. It can be understood as the choice probability derived from a "limiting game" with infinite number of players, where each individual i forms a link with j following a limiting strategy given by

$$G_{ij} = 1\left\{V_{ij}(\sigma) + 2D_j' H(\sigma)\omega_i(\sigma) - \varepsilon_{ij} \geq 0\right\}$$

In this limiting game, conditional on sufficient statistic $\omega_i(\sigma)$ which summarizes the spillover effects from one's other links due to nonseparable utility, each individual makes link decisions myopically and considers each link as a binary choice independent of her other links. Moreover, while statistic $\omega_{n,i}(X, \varepsilon_i, \sigma)$ in the n-player game may be dependent of ε_{ij}, the dependence vanishes as $n \to \infty$ and in the limit statistic $\omega_i(\sigma)$ is independent of ε_{ij}.

Under Assumption 5 that ensures the maximizer of $\Pi(\omega, X_i, \sigma)$ is identified, we can show that the statistic $\omega_{n,i}$ and choice probability $P_n(X_i, X_j; X, \sigma)$ in the finite game converge to the proposed limits given in (17) and (16) as $n \to \infty$.

Assumption 5 *For any* X_i *and symmetric* σ, $\Pi(\omega, X_i, \sigma)$ *defined in (17) has a unique maximizer* $H(\sigma)\omega_i^*(\sigma)$.

Proposition 3 *Under Assumptions 1-5, given any* X_i,

$$\sup_{\sigma} |H(\sigma)\omega_{n,i}(X, \varepsilon_i, \sigma) - H(\sigma)\omega_i^*(\sigma)| \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{18}$$

Moreover, given any (X_i, X_j), *we have*

$$\sup_{\sigma} |P_n(X_i, X_j; X, \sigma) - P(X_i, X_j; \sigma)| \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{19}$$

Proof. See the appendix. ■

Remark 1 *In Proposition 3 we state the convergence of* $\omega_{n,i}$ *to* ω_i *in terms of* $H\omega_{n,i}$ *and* $H\omega_i$ *because* H *may be singular so* $\omega_{n,i}$ *and* ω_i *may not be unique. Nevertheless, note that only* $H\omega$ *and* $\omega' H\omega$ *enter (14)-(15) and (16)-(17). If two* $\omega_1 \neq \omega_2$ *satisfy* $H\omega_1 = H\omega_2$, *then* $\omega_1' H\omega_1 - \omega_2' H\omega_2 = (\omega_1 + \omega_2)' H(\omega_1 - \omega_2) = 0$ *where the second equality is by symmetry of* H. *Hence it is enough to consider* $H\omega$.

5 Estimation

In this section, we discuss the estimation of model parameter θ. We assume that a single large network is observed[3] and propose a simple two step estimator. Unlike the standard literature on incomplete information games which considers many-game asymptotics, we face the challenge that an n-dependent link formation probability needs to be estimated from a single game. Our idea is to use the symmetry feature of an equilibrium, which ensures links between individuals with given characteristics are identically distributed. This idea of doing asymptotics based on symmetry was used by Leung (2015) as well. We extend it to the case of nonseparable utility. In our case, links formed by different individuals are independent conditional on observables due to independent private shocks, while links formed by one individual are dependent through statistic $\omega_{n,i}(X, \varepsilon_i)$ which converges to a constant ω_i as n approaches infinity. Such vanishing dependence across links as n increases indicates that asymptotics for large n can be valid.

To fix ideas, we restrict the specification of separable utility $v_{ij}(G_{-i}, X)$ to be also separable in other players. That is, $v_{ij}(G_{-i}, X)$ depends on G_{-i} through a linear combination

[3] If more than one network is observed, we can proceed network by network. That is, we estimate link formation probabilities network by network and pool likelihoods or moments from each network to estimate θ.

of terms $G_{i_1 j_1} G_{i_2 j_2} \cdots G_{i_k j_k}$ such that $i_1, i_2, \ldots, i_k \in \mathcal{N}_n \backslash \{i\}$ are distinct. This restriction ensures that the expected utility in (4) only involves the probability of forming a link as in (15), so it is enough to propose estimators for such one-link formation probabilities. We see later in Section 6 that in undirected networks the expected utility may involve the probability of forming two links, so there we also need estimators for those two-link formation probabilities.

For notational simplicity, let $X_{ij} = (X_i, X_j)$ and $x^{st} = (x^s, x^t) \in \mathcal{X}^2$, $1 \leq s, t \leq T$. Define $p_{n,st}(X, \sigma)$ to be the probability that a pair with characteristics x^{st} forms a link conditional on X and σ

$$p_{n,st}(X, \sigma) = \Pr\left(G_{n,ij} = 1 | X_{ij} = x^{st}, X, \sigma\right)$$

It can be estimated by the empirical frequency that such pairs form a link

$$\hat{p}_{n,st} = \frac{\sum_i \sum_{j \neq i} G_{n,ij} 1\{X_{ij} = x^{st}\}}{\sum_i \sum_{j \neq i} 1\{X_{ij} = x^{st}\}}$$

which we refer to as the link frequency estimator. The following proposition shows that the link frequency estimator is consistent.

Proposition 4 *For any X and symmetric σ,*

$$\sup_{s,t} |\hat{p}_{n,st} - p_{n,st}(X, \sigma)| = o_p(1)$$

as $n \to \infty$.

Proof. See the appendix. ■

With the first stage estimates $\hat{p}_n = (\hat{p}_{n,st})_{s,t \leq T}$, we can estimate θ by pseudo MLE of GMM. In the former, we construct the pseudo log likelihood

$$\ln \mathcal{L}_n(\theta, p_n) = \sum_i \sum_{j \neq i} G_{ij} \ln P_n(X_{ij}; X, p_n, \theta) + (1 - G_{ij}) \ln(1 - P_n(X_{ij}; X, p_n, \theta)) \quad (20)$$

A maximum likelihood estimator of θ is given by

$$\hat{\theta}_n = \arg\max_\theta \ln \mathcal{L}_n(\theta, \hat{p}_n)$$

For GMM, define population moment

$$m_{st}(\theta, p_n) = \mathbb{E}\left[G_{n,ij} - P_n(X_{ij}; X, p_n, \theta) | X_{ij} = x^{st}, X, p_n\right]$$

for $1 \leq s, t \leq T$ and $m(\theta, p_n) = (m_{st}(\theta, p_n))_{s,t \leq T}$. The model implies $m(\theta, p_n) = 0$. We can estimate the population moment $m_{st}(\theta, p_n)$ by

$$m_{n,st}(\theta, \hat{p}_n) = \frac{\sum_i \sum_{j \neq i} (G_{n,ij} - P_n(X_{ij}; X, \hat{p}_n, \theta)) 1\{X_{ij} = x^{st}\}}{\sum_i \sum_{j \neq i} 1\{X_{ij} = x^{st}\}}$$

and stack them into a vector $m_n(\theta, \hat{p}_n) = (m_{n,st}(\theta, \hat{p}_n))_{s,t \leq T}$. The GMM estimator of θ is given by

$$\hat{\theta}_n = \arg\min_{\theta} m_n(\theta, \hat{p}_n)' W_n(\theta) m_n(\theta, \hat{p}_n)$$

with $W_n(\theta)$ a weighting matrix.

6 Undirected Networks

In this section we consider the case of undirected networks. We show that the idea in previous sections could also work for undirected networks with mild modifications. To proceed, let G_{ij} denote an undirected link between i and j. Clearly $G_{ij} = G_{ji}$. It is useful to denote by S_{ij} a directed link from i to j. Under the link announcement framework, S_{ij} represents whether i proposes to form a link with j. The link is formed if both i and j propose to form it, so $G_{ij} = S_{ij} S_{ji}$. We may write $G(S)$ to indicate that G is the network induced by proposals S.

We consider the same utility specification as in (1), with G_{ij} representing an undirected link. With abuse of notation we use G_{-i} to denote the submatrix of G_{-i} with the ith row and column deleted. We maintain the same information assumption. The strategy of individual i, denoted by $S_i(X, \varepsilon_i) = (S_{ij}(X, \varepsilon_i))_{j \in \mathcal{N}_n}$ with $S_{ii} = 0$, is a mapping from her information (X, ε_i) to a vector of link proposals in $\mathcal{S}_i = \{0, 1\}^{n-1}$. The optimal strategy maximizes her expected utility $\mathbb{E}[U_i(G(S_i, S_{-i}), X, \varepsilon_i) | X, \varepsilon_i, \sigma]$ over $s_i \in \mathcal{S}_i$, where the expectation is taken with respect to others' proposals $S_{-i} = (S_j)_{j \neq i}$. The choice probability $\sigma_i(s_i | X)$ and belief profile $\sigma(X)$ are defined similarly as before, with links G_i replaced by proposals S_i. Given belief profile σ. individual i's expected utility is now given by

$$\begin{aligned}
&\mathbb{E}[U_i(G(S_i, S_{-i}), X, \varepsilon_i) | X, \varepsilon_i, \sigma] \\
&= \frac{1}{n-1} \sum_{j \neq i} S_{ij} \left(\sum_{s_{-i}} s_{ji} v_{ij}(g_{-i}(s_{-i}), X) \prod_{j \neq i} \sigma_j(s_j | X) - \sigma_{ji} \varepsilon_{ij} \right) \\
&+ \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} S_{ij} S_{ik} \left(\sum_{s_{-i}} s_{ji} s_{ki} h_i(g_j(s_{-i}), g_k(s_{-i})) \prod_{j \neq i} \sigma_j(s_j | X) \right) \qquad (21)
\end{aligned}$$

Let $\Pr\left(S_i = s_i | X, \sigma\right)$ be the probability that i proposes s_i given X and σ. A Bayesian Nash equilibrium $\sigma^*(X)$ is a fixed point that solves $\sigma_i^*\left(s_i | X\right) = \Pr\left(S_i = s_i | X, \sigma^*(X)\right)$.

Remark 2 *A potential concern with Nash is that in undirected networks players may coordinate. This is reasonable under complete information, where pairwise stability (Jackson and Wolinsky (1996)) and Nash equilibrium are nonnested and neither of them implies the other. However, under incomplete information players won't be able to coordinate even in undirected networks; because i does not observe ε_{ji}, he cannot predict what j proposes and coordinate on that (unless in a trivial equilibrium where $\sigma(X) \equiv 0$). In fact, if we define a Bayesian version of the pairwise stability, that is, a network G is Bayesian pairwise stable if*

$$G_{ij} = 1\left\{\Delta_{ij}\mathbb{E}\left[U_i(G\left(S_i, S_{-i}\right), X, \varepsilon_i) | X, \varepsilon_i, \sigma\right] \geq 0 \ \& \ \Delta_{ji}\mathbb{E}\left[U_j(G\left(S_j, S_{-j}\right), X, \varepsilon_j) | X, \varepsilon_j, \sigma\right] \geq 0\right\}$$

for any $i \neq j$, where $\Delta_{ij}\mathbb{E}\left[U_i(G\left(S_i, S_{-i}\right), X, \varepsilon_i) | X, \varepsilon_i, \sigma\right]$ is the expected marginal utility if i proposes the link with j, i.e.,

$$\mathbb{E}\left[U_i(G\left(S_i : S_{ij} = 1, S_{-i}\right), X, \varepsilon_i) | X, \varepsilon_i, \sigma\right] - \mathbb{E}\left[U_i(G\left(S_i : S_{ij} = 0, S_{-i}\right), X, \varepsilon_i) | X, \varepsilon_i, \sigma\right]$$

and similar for $\Delta_{ji}\mathbb{E}\left[U_j(G\left(S_j, S_{-j}\right), X, \varepsilon_j) | X, \varepsilon_j, \sigma\right]$, then any undirected network that is Bayesian Nash must also be Bayesian pairwise stable. This is because for a Bayesian Nash G, $G_{ij} = 1$ if and only if $S_{ij} = S_{ji} = 1$ are optimal, so the expected marginal utility from the link must be nonnegative for both i and j. It is thus enough to consider Bayesian Nash equilibrium.

Unlike in the directed case, the quadratic form in the expected utility (21) may depend on X_i, through the proposals S_{ji} and S_{ki} made to i. Nevertheless, it is still symmetric in j and k, so the eigendecomposition and Legendre transform still apply. Define

$$\begin{aligned} H_{i,st}\left(\sigma\left(X_{-jk}\right)\right) &= \mathbb{E}\left[S_{ji}S_{ki}h_i\left(G_j, G_k\right) | \sigma\left(X_j = x^s, X_k = x^t, X_i, X_{-ijk}\right)\right] \\ &= \sum_{s_{-i}} s_{ji}s_{ki}h_i\left(g_j\left(s_{-i}\right), g_k\left(s_{-i}\right)\right) \prod_{j \neq i} \sigma_j\left(s_j | X_j = x^s, X_k = x^t, X_i, X_{-ijk}\right) \end{aligned}$$

and the matrix

$$H_i\left(\sigma\left(X\right)\right) = \begin{bmatrix} H_{i,11}\left(\sigma\left(X\right)\right) & \cdots & H_{i,1T}\left(\sigma\left(X\right)\right) \\ \vdots & & \vdots \\ H_{i,T1}\left(X_i, \sigma\left(X\right)\right) & \cdots & H_{i,TT}\left(\sigma\left(X\right)\right) \end{bmatrix}$$

where the subscript i indicates that they may depend on X_i. Following the same idea in previous sections, we can derive the optimal strategies in closed form.

Corollary 1 *Under Assumptions 1-3(with $H(\sigma)$ replaced with $H_i(\sigma)$), the optimal strategy $S_{n,i}(X,\varepsilon_i,\sigma) = (S_{n,ij}(X,\varepsilon_i,\sigma))_{j\neq i}$ is given by*

$$S_{n,ij}(X,\varepsilon_i,\sigma) = 1\left\{V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j' H_i(\sigma)\,\omega_{n,i}(X,\varepsilon_i,\sigma) - \sigma_{ji}\varepsilon_{ij} \geq 0\right\},\ \forall j \neq i$$

where $\omega_{n,i}(X,\varepsilon_i,\sigma)$ is a maximizer of the problem

$$\max_{\omega} \frac{1}{n-1}\sum_{j\neq i}\left[V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j' H_i(\sigma)\,\omega - \sigma_{ji}\varepsilon_{ij}\right]_+ - \frac{n-1}{n-2}\omega' H_i(\sigma)\,\omega$$

with

$$V_{n,ij}(X,\sigma) = \sum_{s_{-i}} s_{ji} v_{ij}(g_{-i}(s_{-i}), X)\prod_{j\neq i}\sigma_j(s_j|\,X) - \frac{1}{n-2}D_j' diag(H_i(\sigma))\,D_j$$

Moreover, the optimal $S_{n,ij}(X_i,X_j)$ and $H_i(\sigma)\,\omega_{n,i}(X,\varepsilon_i,\sigma)$ are unique almost surely.

Given the optimal strategies in the corollary, the probability that individual i proposes to link to j and the probability that i proposes to link to both j and k are given by

$$\begin{aligned} P_n(X_i,X_j;X,\sigma) &= \Pr(S_{n,ij}(X,\varepsilon_i,\sigma) = 1|X,\sigma) \\ &= \Pr\left(V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j' H_i(\sigma)\,\omega_{n,i}(X,\varepsilon_i,\sigma) - \sigma_{ji}\varepsilon_{ij} \geq 0|X,\sigma\right) \end{aligned}$$

and

$$\begin{aligned} Q_n(X_i,X_j,X_k;X,\sigma) &= \Pr(S_{n,ij}(X,\varepsilon_i,\sigma) = 1, S_{n,ik}(X,\varepsilon_i,\sigma) = 1|X,\sigma) \\ &= \Pr\Bigg(V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j' H_i(\sigma)\,\omega_{n,i}(X,\varepsilon_i,\sigma) - \sigma_{ji}\varepsilon_{ij} \geq 0\ \& \\ &\qquad V_{n,ik}(X,\sigma) + \frac{n-1}{n-2}2D_k' H_i(\sigma)\,\omega_{n,i}(X,\varepsilon_i,\sigma) - \sigma_{ki}\varepsilon_{ik} \geq 0\,\Bigg|\, X,\sigma\Bigg) \end{aligned}$$

The latter is relevant because the expected utility in (21) indicates that individual i may need to form a belief about $S_{n,ji}S_{n,jk}$. We expect that these choice probabilities converge to limiting probabilities

$$P(X_i,X_j;\sigma) = \Pr\left(V_{ij}(\sigma) + 2D_j' H_i(\sigma)\,\omega_i^*(\sigma) - \sigma_{ji}\varepsilon_{ij} \geq 0|X_i,X_j,\sigma\right)$$

and

$$Q(X_i, X_j, X_k; \sigma) = P(X_i, X_j; \sigma) \cdot P(X_i, X_k; \sigma)$$

respectively as $n \to \infty$, where $\omega_i^*(\sigma)$ is a maximizer of the problem

$$\max_{\omega} \mathbb{E}\left(\left[V_{ij}(\sigma) + 2D_j' H_i(\sigma)\omega - \sigma_{ji}\varepsilon_{ij} \right]_+ \middle| X_i \right) - \omega' H_i(\sigma)\omega$$

and $V_{ij}(\sigma)$ is a limit of $V_{n,ij}(X, \sigma)$ that satisfies Assumption 4. That is,

$$\begin{aligned} \sup_{\sigma} |P_n(X_i, X_j; X, \sigma) - P(X_i, X_j; \sigma)| &= o_p(1) \\ \sup_{\sigma} |Q_n(X_i, X_j, X_k; X, \sigma) - Q(X_i, X_j, X_k; \sigma)| &= o_p(1) \end{aligned}$$

7 Simulation

In this section, we implement the proposed methods in a simulation study. We focus on directed networks and assume the following utility specification

$$\begin{aligned} U_i(G, X, \varepsilon_i; \theta) = & \frac{1}{n-1} \sum_{j \neq i} G_{ij} \left(\beta_0 + X_i \beta_1 + |X_i - X_j| \beta_2 + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk}\beta_3 - \varepsilon_{ij} \right) \\ & + \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} G_{ij} G_{ik} G_{jk} G_{kl} \gamma \end{aligned}$$

where X_i are i.i.d. binary random variables taking values in $\{0, 1\}$ with equal probability and ε_{ij} are i.i.d. standard normal. The true values of the parameters are set at $(\beta_{00}, \beta_{10}, \beta_{20}, \beta_{30}, \gamma_0) = (-1, 1, -2, 1, 1)$. The networks are generated according to the n-player incomplete information game described in Section 2, with n taking values of 10, 25, 50, 100, 250, and 500. For each value of n, we generate a single network and use it to estimate the parameters by two-step MLE or GMM. Each experiment is repeated 100 times. We report the means and standard errors of the estimated parameters.

Since the limiting game approximates the finite game asymptotically, we first use the limiting game to estimate the parameters and check how well such estimates would perform. In particular, we construct the likelihood as in (20), with $P_n(X_{ij}; X, p_n)$ replaced by the limiting choice probability $P(X_{ij}; p_n)$ given in (16). Such an approximation has a substantial advantage in computation because $P(X_{ij}; p_n)$ has a probit-type closed form. The estimates are reported in Table 1. It is not surprising that the estimates in small networks perform poorly, as the limiting game is only an asymptotic approximation of the finite game. Nevertheless, when the network size gets large (e.g. $n \geq 100$), the estimates become close to the truth. This indicates that the limiting game is a valid approximation of the finite game asymptotically and estimation based on this approximation may yield good estimates for the parameters if networks are sufficiently large.

Next we estimate the parameters by the finite game, and compare the estimates with those from the limiting game. Note that the finite-game choice probability $P_n(X_{ij}; X, p_n)$ does not have a closed form because $\omega_{n,i}(X, \varepsilon_i)$ depends on ε_i. It needs to be computed by simulation. In practice, we simulate $P_n(X_{ij}; X, p_n)$ by a frequency simulator, where in each simulation we generate a vector ε_i and for this ε_i we solve for the optimal $G_{n,ij}(X, \varepsilon_i)$ numerically. More specifically, we obtain $G_{n,ij}(X, \varepsilon_i)$ by either solving the integer programming problem itself (i.e., maximizing the expected utility) when $n \leq 100$ or applying the equivalent binary representation as in (13) and solving for $\omega_{n,i}(X, \varepsilon_i)$ when $n > 100$. Table 2 reports the MLE estimates using the finite game. By correctly specifying the choice probability, we improve the estimates in small networks. Using the finite game also improves the estimation precision. The standard errors of the estimators are smaller than those from the limiting game, for all the parameters and all n. These results suggest us to use the finite game rather than the limiting game when n is relatively small, which is reasonable given that the computational gain of the limiting approximation is substantial only for large n.

Because the simulated maximum likelihood estimators could be biased unless the number of simulations is infinite due to nonlinearity of the likelihood, we also consider GMM estimators where we use the optimal weights derived from the score of the likelihood. We simulate the weights independently from the simulated choice probabilities, using either the finite game or the limiting game, to avoid a finite-simulation bias. The estimates are reported in Tables 3 and 4. As we have expected, the GMM estimates with both the choice probabilities and weights simulated using the finite game (Table 3) have a smaller bias than the MLE estimates for small n. Note that the GMM estimation takes about twice of the computational time of the MLE, because the choice probabilities and weights need to be simulated independently. If instead we approximate the weights using the limiting game, the estimates (Table 4) outperform the MLE estimates for $n = 25$, but lead to a bigger bias for $n = 10$. Overall, GMM based entirely on the finite game provides the least biased estimates and is thus recommended for small n.

8 Conclusions

In this paper, we provide estimation methods for network formation using observed data from a single large network. We model network formation as a simultaneous-move game with private information and extend Leung (2015) by allowing for nonlinear utility such as

Table 1: MLE Estimates Using the Limiting Game

n	β_0	β_1	β_2	β_3	γ
10	−1.152	2.806	−6.469	−2.626	−0.194
	(2.284)	(3.004)	(3.649)	(8.890)	(6.438)
25	−0.719	2.639	−3.899	−1.835	−0.887
	(0.447)	(2.029)	(2.152)	(3.710)	(3.948)
50	−0.986	1.058	−2.064	0.858	0.909
	(0.126)	(0.499)	(0.499)	(0.921)	(0.551)
100	−0.995	1.008	−2.007	0.985	0.959
	(0.034)	(0.084)	(0.084)	(0.165)	(0.208)
250	−1.001	1.004	−2.003	1.009	0.969
	(0.014)	(0.039)	(0.037)	(0.075)	(0.173)
500	−1.001	1.001	−2.000	1.006	0.986
	(0.010)	(0.022)	(0.022)	(0.047)	(0.103)
DGP	−1	1	−2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the limiting game.

Table 2: MLE Estimates Using the Finite Game

n	β_0	β_1	β_2	β_3	γ
10	−1.042	1.143	−2.312	0.901	0.891
	(0.674)	(0.563)	(0.920)	(0.656)	(1.297)
25	−1.005	1.059	−2.123	0.906	0.969
	(0.122)	(0.269)	(0.324)	(0.379)	(0.290)
50	−1.009	0.995	−1.997	1.024	1.011
	(0.063)	(0.154)	(0.148)	(0.186)	(0.221)
100	−0.990	0.992	−2.013	1.006	0.990
	(0.028)	(0.064)	(0.060)	(0.105)	(0.095)
250	−0.995	1.002	−2.004	1.017	0.985
	(0.010)	(0.024)	(0.023)	(0.046)	(0.035)
500	−0.999	1.014	−2.004	0.994	0.982
	(0.006)	(0.014)	(0.014)	(0.027)	(0.022)
DGP	−1	1	−2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the finite game, where the CCPs are computed from 500 simulations by either solving integer programming (for $n \leq 100$) or applying Legendre transform (for $n > 100$).

Table 3: GMM Estimates Using the Finite Game (Weights from the Finite Game)

n	β_0	β_1	β_2	β_3	γ
10	-0.995	0.958	-1.937	0.981	0.979
	(0.206)	(0.191)	(0.402)	(0.185)	(0.182)
25	-1.010	1.060	-2.038	1.003	0.996
	(0.066)	(0.109)	(0.194)	(0.092)	(0.098)
50	-1.003	0.999	-2.001	1.020	0.988
	(0.042)	(0.065)	(0.097)	(0.083)	(0.072)
100	-0.996	0.993	-2.010	1.031	0.981
	(0.023)	(0.036)	(0.052)	(0.064)	(0.055)
250	-0.998	0.999	-2.000	1.027	0.987
	(0.008)	(0.017)	(0.020)	(0.035)	(0.033)
500	-1.001	1.007	-1.997	0.998	0.995
	(0.006)	(0.011)	(0.011)	(0.028)	(0.019)
DGP	-1	1	-2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the finite game, with the GMM weights simulated independently from the finite game. The CCPs are computed from 500 simulations by either solving integer programming (for $n \leq 100$) or applying Legendre transform (for $n > 100$).

Table 4: GMM Estimates Using the Finite Game (Weights from the Limiting Game)

n	β_0	β_1	β_2	β_3	γ
10	-1.008	1.273	-2.940	0.834	0.943
	(0.438)	(1.010)	(3.338)	(1.900)	(1.004)
25	-1.017	1.012	-2.065	1.016	0.986
	(0.097)	(0.185)	(0.268)	(0.232)	(0.146)
50	-1.010	0.995	-1.995	1.050	0.984
	(0.052)	(0.070)	(0.101)	(0.110)	(0.094)
100	-0.995	0.991	-2.010	1.034	0.979
	(0.023)	(0.040)	(0.050)	(0.073)	(0.062)
250	-0.998	1.000	-2.001	1.031	0.983
	(0.008)	(0.018)	(0.021)	(0.038)	(0.036)
500	-1.001	1.010	-2.000	0.999	0.989
	(0.005)	(0.013)	(0.012)	(0.034)	(0.025)
DGP	-1	1	-2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the finite game, with the GMM weights simulated independently from the limiting game. The CCPs are computed from 500 simulations by either solving integer programming (for $n \leq 100$) or applying Legendre transform (for $n > 100$).

the effect of friends in common. The main innovation is to provide an approach to explicitly represent the pure strategy of an individual in the game. This closed form representation enables us to analyze the asymptotic features of the game as the number of players approaches infinity and thus construct asymptotically valid estimators for the parameters. We propose a two-step estimation procedure which makes little assumption about equilibrium selection and is computationally simple. Our approach can apply to both directed and undirected networks. We focus on discrete observables in this paper, but expect our approach could be extended to continuous observables as well.

9 Appendix: Proofs

Proof of Proposition 1. Define the set of symmetric $\sigma(X)$

$$\Sigma^s(X) = \left\{\sigma(X) \in [0,1]^{n2^{n-1}} : \sigma_i(X) = \sigma_j(X) \text{ if } X_i = X_j\right\}$$

It is clear that $\Sigma^s(X)$ is a convex and compact subset of $[0,1]^{n2^{n-1}}$. Equations in (6) forms a mapping from $\Sigma^s(X)$ to $\Sigma^s(X)$, because if $\sigma \in \Sigma^s(X)$, then $\Pr(G_i = g_i | X, \sigma(X)) = \Pr(G_j = g_j | X, \sigma(X))$ for $X_i = X_j$ and $g_i = g_j$ with (g_{ii}, g_{ij}) swapped with (g_{jj}, g_{ji}), so $\Pr(G_i = g_i | X, \sigma(X))$ is also symmetric. The mapping is continuous in σ because the expected utilities are continuous in $\sigma(X)$ and ε_i has a continuous distribution under Assumption 1. By Brouwer's fixed point theorem there is a fixed point. ■

Proof of Proposition 2. It suffices to show the first statement. Denote $D_{it} = 1\{X_i = x^t\}$. The quadratic form in the expected utility is

$$\begin{aligned}
&\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}G_{ik}\mathbb{E}\left[h_i(G_j,G_k)|\,\sigma\right] \\
=\;& \sum_{j\neq i}\sum_{k\neq i} G_{ij}G_{ik}\sum_s\sum_t D_{js}D_{kt}H_{st}(\sigma) - \sum_{j\neq i}G_{ij}\sum_t D_{jt}H_{tt}(\sigma) \\
=\;& \sum_s\sum_t H_{st}(\sigma)\sum_{j\neq i}G_{ij}D_{js}\sum_{k\neq i}G_{ik}D_{kt} - \sum_{j\neq i}G_{ij}\sum_t D_{jt}H_{tt}(\sigma) \\
=\;& \left(\sum_{j\neq i}G_{ij}D_j'\right)H(\sigma)\left(\sum_{j\neq i}G_{ij}D_j\right) - \sum_{j\neq i}G_{ij}D_j' diag\,(H(\sigma))\,D_j \qquad (22)
\end{aligned}$$

By the real spectral decomposition of $H(\sigma)$,

$$\begin{aligned}\left(\sum_{j\neq i} G_{ij}D_j'\right) H(\sigma)\left(\sum_{j\neq i} G_{ij}D_j\right) &= \left(\sum_{j\neq i} G_{ij}D_j'\right)\Phi(\sigma)\Lambda(\sigma)\Phi(\sigma)'\left(\sum_{j\neq i} G_{ij}D_j\right)\\ &= \left(\sum_{j\neq i} G_{ij}D_j'\Phi(\sigma)\right)\Lambda(\sigma)\left(\sum_{j\neq i} G_{ij}\Phi(\sigma)' D_j\right)\\ &= \sum_t \lambda_t(\sigma)\left(\sum_{j\neq i} G_{ij}D_j'\phi_t(\sigma)\right)^2 \qquad (23)\end{aligned}$$

The desired statement then follows from (22), (23), and simple algebra. ■

Proof of Lemma 1. Note that

$$\begin{aligned}&\max_x f(x,y) \geq f(x,y), \quad \forall (x,y)\in \mathcal{X}\times\mathcal{Y}\\ \Rightarrow\ &\max_y \max_x f(x,y) \geq \max_y f(x,y), \quad \forall x\in\mathcal{X}\\ \Rightarrow\ &\max_y \max_x f(x,y) \geq \max_x \max_y f(x,y)\end{aligned}$$

Similarly,

$$\max_x \max_y f(x,y) \geq \max_y \max_x f(x,y)$$

Hence (11) is proved. If (x^*, y^*) is the unique solution to the LHS of (11), then by $f(x^*, y^*) = \max_y \max_x f(x,y) = \max_x \max_y f(x,y)$ it is also the unique solution to the RHS of (11). ■

Proof of Example 2. We verify that $V_{n,ij}(X,\sigma)$ and $V_{ij}(\sigma)$ given in Example 2 satisfy Assumption 4. Note that $\sup_\sigma \left|\frac{1}{n-2}D_j' diag(H(\sigma)) D_j\right| = o_p(1)$ (because $H(\sigma)$ is uniformly bounded) and that $\sup_\sigma \frac{1}{n-1}\sum_{j\neq i}\left|\frac{1}{n-2}D_j' diag(H(\sigma)) D_j\right| = \frac{1}{n-2}\sup_\sigma \frac{1}{n-1}\sum_{j\neq i}\left|D_j' diag(H(\sigma)) D_j\right| = o_p(1)$ by applying uniform law of large numbers to the second sup term, which is appropriate because the space of symmetric σ is compact, $H(\sigma)$ is continuous in σ and is uniformly bounded. It suffices to show that for any θ,

$$\sup_\sigma \left|\frac{1}{n-2}\sum_{k\neq i,j}\sigma_{jk}(X_j,X_k)\, w(X_i,X_k)\,\beta_4 - \mathbb{E}\left[\sigma_{jk}(X_j,X_k)\, w(X_i,X_k)\middle| X_i, X_j\right]\beta_4\right| = o_p(1)$$

and

$$\sup_\sigma \frac{1}{(n-1)}\sum_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\sigma_{jk}(X_j,X_k)\, w(X_i,X_k)\,\beta_4 - \mathbb{E}\left[\sigma_{jk}(X_j,X_k)\, w(X_i,X_k)\middle| X_i, X_j\right]\beta_4\right| = o_p(1)$$

The former is satisfied by applying uniform law of large numbers again. As for the latter, write $\Delta_i(X_j, X_k; \sigma) = \sigma_{jk}(X_j, X_k) w(X_i, X_k) \beta_4 - \mathbb{E}[\sigma_{jk}(X_j, X_k) w(X_i, X_k) | X_i, X_j] \beta_4$, which has zero conditional mean $\mathbb{E}[\Delta_i(X_j, X_k; \sigma) | X_i] = 0$. By Cauchy-Schwarz inequality, we have $\left(\frac{1}{n}\sum_i |y_i|\right)^2 \leq \frac{1}{n}\sum_i y_i^2$, so

$$\begin{aligned}
& \left(\frac{1}{(n-1)(n-2)} \sum_{j \neq i} \left| \sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma) \right| \right)^2 \\
\leq \; & \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \left(\sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma) \right)^2 \\
= \; & \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma)^2 \\
& + \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \Delta_i(X_j, X_k; \sigma) \Delta_i(X_j, X_l; \sigma)
\end{aligned}$$

The last two terms are U-processes. It remains to show that they are $o_p(1)$ uniformly in σ. For the first term, applying Corollary 7 in Sherman (1994) yields

$$\sup_\sigma \left| \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma)^2 - \mathbb{E}\left[\Delta_i(X_j, X_k; \sigma)^2 \middle| X_i \right] \right| = O_p\left(\frac{1}{\sqrt{n}} \right)$$

so the first terms is $O_p\left(\frac{1}{n}\right)$ uniformly in σ. As for the second term, note that the product $\Delta_i(X_j, X_k; \sigma) \Delta_i(X_j, X_l; \sigma)$ has zero mean conditional on X_i, so by Corollary 7 in Sherman (1994) again we obtain

$$\sup_\sigma \left| \frac{1}{(n-1)(n-2)(n-3)} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \Delta_i(X_j, X_k; \sigma) \Delta_i(X_j, X_l; \sigma) \right| = O_p\left(\frac{1}{\sqrt{n}} \right)$$

which proves that the second term is $o_p(1)$ uniformly in σ. The proof is complete. ■

Proof of Proposition 3. Step 1: We first prove (18). Because $\frac{\partial}{\partial c}\mathbb{E}[c-\varepsilon]_+ = \frac{\partial}{\partial c}\int_{-\infty}^{c} (c-\varepsilon) f_\varepsilon(\varepsilon) d\varepsilon = F_\varepsilon(c)$, the first order condition of the problem (17) is

$$\nabla_\omega \Pi(\omega, X_i, \sigma) = 2H(\sigma) \mathbb{E}\left(D_j F_\varepsilon\left(V_{ij}(\sigma) + 2D_j' H(\sigma) \omega \right) \middle| X_i \right) - 2H(\sigma)\omega = 0 \qquad (24)$$

It is easy to see that any $\omega_i(\sigma)$ that satisfies the first order condition must be bounded. Hence without loss of generality we can assume that $\omega_i(\sigma)$ is in a compact set Ω. Since $\Pi(\omega, X_i, \sigma)$ is continuous in σ, the compactness of Ω implies that the unique maximizer

$H(\sigma)\omega_i^*(\sigma)$ (by Assumption 5) is also well separated. If we can further show that

$$\sup_{\omega,\sigma}|\Pi_n(\omega, X, \varepsilon_i, \sigma) - \Pi(\omega, X_i, \sigma)| = o_p(1) \tag{25}$$

then following a standard proof for uniform consistency we can prove (18). Specifically, from (25) we have $\sup_\sigma \Pi_n(\omega_{n,i}, X, \varepsilon_i, \sigma) \geq \sup_\sigma \Pi_n(\omega_i^*, X, \varepsilon_i, \sigma) \geq \sup_\sigma \Pi(\omega_i^*, X_i, \sigma) - o_p(1)$, whence,

$$\begin{aligned}\sup_\sigma \Pi(\omega_i^*, X_i, \sigma) - \sup_\sigma \Pi(\omega_{n,i}, X_i, \sigma) &\leq \sup_\sigma \Pi_n(\omega_{n,i}, X_i, \varepsilon_i, \sigma) - \sup_\sigma \Pi(\omega_{n,i}, X_i, \sigma) + o_p(1) \\ &\leq \sup_{\omega,\sigma}|\Pi_n(\omega, X_i, \varepsilon_i, \sigma) - \Pi(\omega, X_i, \sigma)| + o_p(1) = o_p(1).\end{aligned}$$

Well-separateness of $H(\sigma)\omega_i^*(\sigma)$ implies that for any $\varepsilon > 0$, there is $\eta > 0$ such that, for any symmetric σ, $\Pi(\omega, X_i, \sigma) < \Pi(\omega_i^*, X_i, \sigma) - \eta$ for every ω with $\|H(\sigma)\omega - H(\sigma)\omega_i^*(\sigma)\| \geq \varepsilon$. Therefore,

$$\begin{aligned}\Pr\left(\sup_\sigma \|H(\sigma)\omega_{n,i}(\sigma) - H(\sigma)\omega_i^*(\sigma)\| \geq \varepsilon\right) &\leq \Pr\left(\sup_\sigma [\Pi(\omega_{n,i}, X_i, \sigma) - \Pi(\omega_i^*, X_i, \sigma)] < -\eta\right) \\ &\leq \Pr\left(\sup_\sigma \Pi(\omega_{n,i}, X_i, \sigma) - \sup_\sigma \Pi(\omega^*, X_i, \sigma) < -\eta\right) \\ &\to 0\end{aligned}$$

in view of the preceding display (18) is proved.

Now we prove (25). The left hand side of (25) equals

$$\begin{aligned}&\sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[V_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+ - \mathbb{E}\left(\left[V_{ij}(\sigma) + 2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right)\right| \\ \leq\ &\sup_{\omega,\sigma}\frac{1}{n-1}\sum_{j\neq i}\left[\frac{1}{n-2}2D_j'H(\sigma)\omega\right]_+ \\ &+\sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[V_{n,ij}(X,\sigma) + 2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+ - \left[V_{ij}(\sigma) + 2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+\right| \\ &+\sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[V_{ij}(\sigma) + 2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+ - \mathbb{E}\left(\left[V_{ij}(\sigma) + 2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right)\right|\end{aligned} \tag{26}$$

For the third term in (26), because $\left[V_{ij}(\sigma) + 2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+$ are i.i.d. conditional on X_i with conditional mean $\mathbb{E}\left(\left[V_{ij}(\sigma) + 2D_j'H(\sigma)\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right)$, are continuous in ω and σ, and are bounded by

$$\left[V_{ij}(\sigma)+2D_j'H(\sigma)\omega-\varepsilon_{ij}\right]_+\leq\left[\sup_{\omega,\sigma}\left(V_{ij}(\sigma)+2D_j'H(\sigma)\omega\right)-\varepsilon_{ij}\right]_+$$

which is absolute integrable because of the continuity of $V_{ij}(\sigma)$ and $H(\sigma)$ and compactness of the spaces of ω and σ, uniform law of large numbers holds, so

$$\sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[V_{ij}(\sigma)+2D_j'H(\sigma)\omega-\varepsilon_{ij}\right]_+-\mathbb{E}\left(\left[V_{ij}(\sigma)+2D_j'H(\sigma)\omega-\varepsilon_{ij}\right]_+\middle|X_i\right)\right|=o_p(1)$$

As for the second term, because $\left|[x]_+-[y]_+\right|\leq|x-y|$, we have

$$\begin{aligned}&\sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[V_{n,ij}(\sigma)+2D_j'H(\sigma)\omega-\varepsilon_{ij}\right]_+-\left[V_{ij}(\sigma)+2D_j'H(\sigma)\omega-\varepsilon_{ij}\right]_+\right|\\ \leq\;&\sup_{\sigma}\frac{1}{n-1}\sum_{j\neq i}\left|V_{n,ij}(\sigma)-V_{ij}(\sigma)\right|=o_p(1)\end{aligned}$$

by Assumption 4(ii). Finally, the first term in (26) is $o_p(1)$, again by uniform law of large numbers. Hence (25) is proved.

Step 2: Next we prove (19). By definition of P_n and P,

$$\begin{aligned}&\left|P_n(X_i,X_j;X,\sigma)-P(X_i,X_j;\sigma)\right|\\ \leq\;&\int\left|1\left\{V_{n,ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'H(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)\geq\varepsilon_{ij}\right\}-1\left\{V_{ij}(\sigma)+2D_j'H(\sigma)\omega_i^*(\sigma)\geq\varepsilon_{ij}\right\}\right|dF_{\varepsilon_i}(\varepsilon_i)\\ \leq\;&\Pr\left(V_{n,ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'H(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)\geq\varepsilon_{ij}>V_{ij}(\sigma)+2D_j'H(\sigma)\omega_i^*(\sigma)\middle|X\right)\\ &+\Pr\left(V_{n,ij}(X,\sigma)+2D_j'H(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)<\varepsilon_{ij}\leq V_{ij}(\sigma)+2D_j'H(\sigma)\omega_i^*(\sigma)\middle|X\right)\end{aligned}\tag{27}$$

Since the last two terms are similar, without loss of generality it suffices to show that the second last term is $o_p(1)$ uniformly in σ. Define

$$\begin{aligned}M_n(X,\varepsilon_i,\sigma)&=V_{n,ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'H(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)\\ M(\sigma)&=V_{ij}(\sigma)+2D_j'H(\sigma)\omega_i^*(\sigma)\end{aligned}$$

Fix $\eta>0$. Because $\varepsilon_{ij}\in(M(\sigma),M_n(X,\sigma)]$ implies that $\varepsilon_{ij}\in(M(\sigma),M(\sigma)+\eta]$ or

$M_n(X,\sigma) > M(\sigma)+\eta$, we can bound the second last term in (27) as follows

$$\begin{aligned}
& \Pr\left(M_n(X,\varepsilon_i,\sigma) \geq \varepsilon_{ij} > M(\sigma) \middle| X\right) \\
\leq\ & \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma) \middle| X\right) + \Pr\left(M_n(X,\varepsilon_i,\sigma) > M(\sigma)+\eta \middle| X\right) \\
\leq\ & \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma) \middle| X\right) + \Pr\left(2D_j' H(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma) - \omega_i^*(\sigma)\right) > \frac{\eta}{2} \middle| X\right) \\
& +1\left\{V_{n,ij}(X,\sigma) - V_{ij}(\sigma) > \frac{\eta}{2}\right\}
\end{aligned}$$

where the last inequality follows because $M_n(X,\varepsilon_i,\sigma) - M(\sigma) > \eta$ implies that $V_{n,ij}(X,\sigma) - V_{ij}(\sigma) > \frac{\eta}{2}$ or $2D_j'H(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma) - \omega_i^*(\sigma)\right) > \frac{\eta}{2}$. It suffices to show that the last three terms in the display are $o_p(1)$ uniformly in σ. For any $\delta > 0$, the last term satisfies

$$\begin{aligned}
& \Pr\left(\sup_\sigma 1\left\{V_{n,ij}(X,\sigma) - V_{ij}(\sigma) > \frac{\eta}{2}\right\} > \delta \middle| X_i, X_j\right) \\
\leq\ & \Pr\left(\sup_\sigma \left(V_{n,ij}(X,\sigma) - V_{ij}(\sigma)\right) > \frac{\eta}{2} \middle| X_i, X_j\right) \to 0
\end{aligned}$$

as $n \to \infty$ by Assumption 4(i), so it is $o_p(1)$ uniformly in σ. For the second term, we have

$$\begin{aligned}
& \Pr\left(\sup_\sigma \Pr\left(2D_j'H(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma) - \omega_i^*(\sigma)\right) > \frac{\eta}{2} \middle| X\right) > \delta \middle| X_i, X_j\right) \\
\leq\ & \frac{1}{\delta}\mathbb{E}\left(\sup_\sigma \Pr\left(2D_j'H(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma) - \omega_i^*(\sigma)\right) > \frac{\eta}{2} \middle| X\right) \middle| X_i, X_j\right) \\
\leq\ & \frac{1}{\delta}\mathbb{E}\left(\Pr\left(\sup_\sigma 2D_j'H(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma) - \omega_i^*(\sigma)\right) > \frac{\eta}{2} \middle| X\right) \middle| X_i, X_j\right) \\
=\ & \frac{1}{\delta}\Pr\left(\sup_\sigma 2D_j'H(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma) - \omega_i^*(\sigma)\right) > \frac{\eta}{2} \middle| X_i, X_j\right) \to 0
\end{aligned}$$

as $n \to \infty$ by Markov inequality, law of iterated expectation, and uniform consistency of $H(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)$ proved earlier, so the second term is also $o_p(1)$ uniformly in σ. As for the first term,

$$\begin{aligned}
& \sup_\sigma \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma) \middle| X\right) \\
=\ & \sup_\sigma \left(F_{\varepsilon_{ij}}(M(\sigma)+\eta) - F_{\varepsilon_{ij}}(M(\sigma))\right) \\
=\ & \sup_\sigma f_{\varepsilon_{ij}}(M(\sigma)+\tilde{\eta})\eta
\end{aligned}$$

for some $\tilde{\eta} \in [0,\eta]$ where the last equality is from mean value theorem. Since η is arbitrary, choosing η to be $o(1)$, we can get $\sup_\sigma \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma) \middle| X,\sigma\right) = o_p(1)$. The proof is complete. ■

Proof of Proposition 4.

$$\begin{aligned}
&\sup_{s,t} \left|\hat{p}_{n,st} - p_{n,st}(X,\sigma)\right| \\
= &\sup_{s,t} \left| \frac{\sum_i \sum_{j\neq i} \left(G_{n,ij} - \Pr\left(G_{n,ij}=1 \mid X_{ij}=x^{st}, X, \sigma\right)\right) 1\{X_{ij}=x^{st}\}}{\sum_i \sum_{j\neq i} 1\{X_{ij}=x^{st}\}} \right|
\end{aligned}$$

Denote the fraction term by $\Delta_{n,st}$. It suffices to show for any $\delta > 0$,

$$\lim_{n\to\infty} \Pr\left(\sup_{s,t} |\Delta_{n,st}| > \delta\right) = 0.$$

By law of iterated expectation and dominated convergence theorem it suffices to show

$$\lim_{n\to\infty} \Pr\left(\sup_{s,t} |\Delta_{n,st}| > \delta \,\middle|\, X, \sigma\right) = o_p(1)$$

Note

$$\begin{aligned}
\Pr\left(\sup_{s,t} |\Delta_{n,st}| > \delta \,\middle|\, X, \sigma\right) &\leq \sum_{s,t} \Pr\left(|\Delta_{n,st}| > \delta \mid X, \sigma\right) \\
&\leq \sum_{s,t} \frac{\mathbb{E}\left(\Delta_{n,st}^2 \mid X, \sigma\right)}{\delta^2} \\
&\leq \frac{T^2}{\delta^2} \max_{s,t} \mathbb{E}\left(\Delta_{n,st}^2 \mid X, \sigma\right)
\end{aligned}$$

where the second inequality is by (conditional version of) Chebyshev's inequality and $\mathbb{E}\left[\Delta_{n,st} \mid X, \sigma\right] = 0$. It suffices to show $\mathbb{E}\left(\Delta_{n,st}^2 \mid X, \sigma\right) = o_p(1)$.

$$\begin{aligned}
&\mathbb{E}\left(\Delta_{n,st}^2 \mid X, \sigma\right) \\
= &\frac{\frac{1}{n^2(n-1)^2} \sum_i \sum_{j\neq i} Var\left(G_{n,ij} \mid X_{ij}=x^{st}, X, \sigma\right) 1\{X_{ij}=x^{st}\}}{\left(\frac{1}{n(n-1)} \sum_i \sum_{j\neq i} 1\{X_{ij}=x^{st}\}\right)^2} \\
&+ \frac{\frac{1}{n^2(n-1)^2} \sum_i \sum_{j\neq i} \sum_{k\neq i,j} Cov\left(G_{n,ij}, G_{n,ik} \mid X_{ij}=X_{ik}=x^{st}, X, \sigma\right) 1\{X_{ij}=X_{ik}=x^{st}\}}{\left(\frac{1}{n(n-1)} \sum_i \sum_{j\neq i} 1\{X_{ij}=x^{st}\}\right)^2}
\end{aligned}$$

where we have used the fact that links formed by different individuals are independent, i.e.,

$Cov\left(G_{n,ij}, G_{n,i'j'}\right| X_{ij} = X_{i'j'} = x^{st}, X, \sigma) = 0$ for all $i \neq i'$. The first term is at most

$$\frac{1}{4n\left(n-1\right)}\left(\frac{1}{n\left(n-1\right)}\sum_{i}\sum_{j\neq i}1\left\{X_{ij}=x^{st}\right\}\right)^{-1}$$

As for the second term, because $Cov\left(G_{n,ij}, G_{n,ik}\right| X_{ij} = X_{ik} = x^{st}, X, \sigma) \leq Var\left(G_{n,ij}\right| X_{ij} = x^{st}, X, \sigma)^{1/2} \cdot Var\left(G_{n,ik}\right| X_{ik} = x^{st}, X, \sigma)^{1/2} \leq \frac{1}{4}$, so the second term is bounded by

$$\frac{1}{4n}\left(\frac{1}{n\left(n-1\right)\left(n-2\right)}\sum_{i}\sum_{j\neq i}\sum_{k\neq i,j}1\left\{X_{ij}=X_{ik}=x^{st}\right\}\right)\left(\frac{1}{n\left(n-1\right)}\sum_{i}\sum_{j\neq i}1\left\{X_{ij}=x^{st}\right\}\right)^{-2}$$

The double and triple summations in the last two displays are U-statistics. It is easy to show that they converge to their expectations. Therefore, the sum of the two displayed terms are $o_p\left(1\right)$, as desired. The proof is complete. ■

References

[1] Aradillas-Lopez, A. (2010) Semiparametric estimation of a simultaneous game with incomplete infomraiton. *Journal of Econometrics*, 157, 409-431.

[2] Bajari, P. Hong, H., Krainer, J., & Nekipelov, D. (2010) Computing equilibria in static games of incomplete information using the all-solution homotopy, mimeo, University of Minnesota.

[3] Bajari, P., Hahn, J., Hong, H., & Ridder, G. (2011). A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models. *International Economic Review*, 53(3), 807-824.

[4] De Paula, Richards-Shubik and Tamer (2015) Identification of Preferences in Network Formation Games, working paper.

[5] Dube, J., Fox, J. T., & C. Su (2012) Improving the numerical performance of static and dynamic aggregated discrete choice random coefficients demand estimation. *Econometrica*, 80(5), 2231-2267.

[6] Jackson, M. O., & Wolinsky, A. (1996) A strategic model of social and economic networks. *Journal of Economic Theory*, 71, 44-74.

[7] Leung, M. (2015) Two-Step Estimation of Network-Formation Models with Incomplete Information, *Journal of Econometrics*, 188: 182-195.

[8] Miyauchi (2013) Structural Estimation of a Pairwise Stable Network with Nonnegative Externality, working paper.

[9] Menzel, K. (2016a) Inference for Games with Many Players, *Review of Economic Studies,* 83(1): 306-337.

[10] Menzel, K. (2016b) Strategic Network Formation with Many Players, NYU, working paper.

[11] Sheng, S. (2016) A Structural Econometric Analysis of Network Formation Games, UCLA, working paper.

[12] Sherman, R. (1994) Maximal Inequalities for Degenerate U-processes with Applications to Optimization Estimators. *Annals of Statistics*, 1994(22), 439-459.

[13] Su, C., & K. L. Judd (2012) Constrained optimization approaches to estimation of structural models. *Econometrica*, 80(5), 2213-2230.