Identification and Estimation of Large Network Games with Private Link Information

Hülya Eraslan and Xun Tang*

October 7, 2023

Abstract

We study the identification and estimation of large network games in a linear-quadratic utility framework, where each individual chooses a continuous action while holding private information about its links *and* payoffs. Extending Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2010), we build a tractable empirical model of network games where the individuals are heterogeneous with private link and payoff information, and characterize its unique, symmetric pure-strategy Bayesian Nash equilibrium. We show that the parameters in individual payoffs are identified under "large market" asymptotics, whereby the number of individuals increases to infinity on a single large network. We also propose a consistent two-step m-estimator for individual payoffs. Our method is distribution-free in that it does not require parametrization of the distribution of shocks in individual payoffs. Monte Carlo simulation shows that our estimator has good performance in moderate-sized samples.

Keywords: Identification, Estimation, Large games, Network games, Private link information

---

*Eraslan: Rice University, Department of Economics, Houston, Texas. E-mail: eraslan@rice.edu. Tang: Rice University, Department of Economics, Houston, Texas. E-mail: xun.tang@rice.edu. We are grateful to Angelo Mele and Matthew Shum for helpful comments. We thank seminar and conference participants at Vanderbilt University and "The Third Annual Conference on Network Science in Economics" and four referees for their helpful comments. İbrahim Emirahmetoḡlu provided excellent research assistance.

1 Introduction

We study the identification and estimation of large Bayesian games on networks where individuals choose continuous actions in a linear-quadratic utility framework, while holding private information about their links as well as payoffs. Private information on links is prevalent in many empirical environments where networks involve large population. In such cases, it is implausible to assume that each individual has complete information about the *full* network structure. For example, Banerjee, Chandrasekhar, Duflo and Jackson (2014) present evidence that members of rural communities in India had incomplete information about the network structure based on the diffusion of gossips. The impact of private information on links in network games has been studied theoretically by Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2010) in their pioneering work. They establish the existence of symmetric, monotone Bayesian Nash equilibria in such games and show that shifts in the degree distribution due to increased connectivity have unambiguous effects on the equilibrium behavior of individuals on the network.

Despite their empirical relevance and theoretical importance, large network games with private links have not been investigated in the structural econometrics literature. Menzel (2015) studies the identification and estimation of large Bayesian games where all individuals are strategically interdependent with exchangeable generic types and actions conditional on observed characteristics, and where private payoff shocks are *conditionally independent* across players. In our model, individual types consist of their links and are therefore inherently correlated between neighbors, thus inducing dependence between their actions in equilibrium.

We provide a new econometric framework for the structural analysis of large network games with private information on links and payoffs. We extend Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2010) to build a tractable empirical model of network games where the individuals are heterogeneous with such private information.[1] Under a quadratic payoff specification (common in social interaction models), we show that the model has a

[1]The model in Galeotti et al (2010) focuses on homogeneous individuals with private information only about their degrees. Each individual's payoff depends on its degree and neighbors' actions. In comparison, our model accommodates a second source of private information in payoff shocks. This is because we consider data environments where the links are reported in the data, and thus need to allow for unreported private information to avoid degeneracy in the econometric model.

unique, symmetric Bayesian Nash equilibrium. We then show the model is identified under the concept of "large-market" asymptotics, whereby the number of networks is fixed and small but the sample size increases with the number of individuals. Such an asymptotic concept has been used in the estimation of large Bayesian games in different contexts. See, for example, Leung (2015), Menzel (2015), Lin and Xu (2017), Kojevnikov and Song (2023) and Xu (2018).[2]

Identification under such a concept for asymptotics is non-standard, because the sample units cannot be considered as independent draws from any "population" that can be treated as known as the sample size approaches infinity. Our identification strategy tackles this difficulty through two steps. First, we define a set of asymptotic moments that can be consistently estimated via sample analogs under the "large-market" asymptotics. Next, we derive the structural link between these asymptotic moments and the model parameters, and use it as the basis for our identification. We then introduce an index sufficiency condition that reduces the dimension of arguments in the individuals' endogenous peer effects, and show that it is sufficient for recovering individual payoff parameters.

We propose a two-step estimator for individual payoffs. The first step estimates the asymptotic moments. The second step proposes m-estimation for the payoff parameters. This estimator is consistent under standard regularity conditions, and has several desirable properties. It does not rely on parametric assumption on payoff shocks and exploits asymptotic uncorrelation between neighborhood profiles; its implementation also does not involve any tuning parameters except those used in the first-step estimation. Monte Carlo simulation shows our estimator has good performance in moderate-sized samples.

A formal characterization of the limiting distribution of our estimator under primitive conditions, as well as construction of standard errors, would require elaborate conditions that restrict the strength of network dependence. Several recent works, e.g., Leung (2021) and Leung and Moon (2023), have tackled these issues of inference, such as normal approximation in the presence of network dependence, in broad contexts. We do not address these challenges of inference in this paper; instead we focus on introducing the framework of identification

[2]Menzel (2015) and Kojevnikov and Song (2023) study large Bayesian games in a general setup where all individuals are strategically interlinked. Leung (2015) estimates large games in network formation when individuals have private payoff shocks, rather than Bayesian games on a given network.

using asymptotic moments, and establishing consistency of the two-step estimator. We do acknowledge that an interesting direction for future research would be to specify primitive conditions in our context that would lead to the form of weak dependence in Leung and Moon (2023).

To reiterate, our empirical strategy is centered around the use of asymptotic moments, i.e., probability limits of average individual types or choices that are interdependent in equilibrium. By leveraging the symmetry and asymptotic uncorrelation between individual units, we show how to use index or exclusion restrictions to recover the underlying payoffs from asymptotic moments. This multi-step approach extends the classical framework for identification with i.i.d. observations, and is generalizable beyond the current setting of network games with private links, continuous choices and a linear quadratic utility.

Our model of private links can be applied in a variety of contexts where individual outcomes are continuous and result from simultaneous strategic interactions within social-economic groups or networks. Todd and Wolpin (2018) estimate a model of the effort decisions by students and teachers in classrooms (groups). Their survey data include continuous measurements of efforts (such as the number of hours per week spent on studying mathematics, the number of days a student skipped a math class, and the percentage of time the student paid attention in class). Giorgi, Frederiksen and Pistaferri (2020) use tax record and matched employer-employee data to define social groups and estimate the peer effects on individual consumption, which has continuous measurement. Broadly speaking, our model can be fitted in similar settings where social groups are known to have a network structure, and where there is need for relaxing the assumption on individual knowledge of network links. We leave this as an interesting direction for future empirical research.

► **Related Literature.** Graham and Hahn (2005) study the identification and estimation of a linear-in-means model of social interactions with unobserved group effects. Lee, Liu and Lin (2010) estimate a social network model with peer, contextual and correlated effects, using network structure as a source of identifying power. In comparison, our model is a large simultaneous game with continuous choices and linear-quadratic utility, where individuals have private knowledge of links as well as payoff shocks. In our case, the structural form for individual choices in equilibrium is linear, but involves each individual's endogenous, rational belief about other individuals' choices. We establish how, in a single large network setting, such beliefs relate to the asymptotic moments, and provide conditions for consistent estimation of these moments using sample averages over the *single* network. These asymptotic moments are then used to point identify the payoff parameters.

Bramoullé, Djebbari and Fortin (2009) and Blume, Brock, Durlauf and Jayaraman (2015) establish identification results in Bayesian games where individuals have private payoff shocks. Both papers maintain that the network structure is common knowledge among all individuals. In addition, the methods proposed in these papers assume knowledge of the reduced-form coefficients for a fixed, unknown network structure. The identification of such reduced-form coefficients per se requires data to contain independent draws of outcomes and explanatory variables generated from that fixed network of interest. To estimate these reduced-form coefficients in practice, one needs a sample that consists of a large number of repeated outcomes from the fixed network.

Xu (2018) and Lin and Xu (2017) estimate discrete-choice Bayesian games on large networks where individuals have private shocks in payoffs, but have common knowledge about the complete network structure. They require "near-epoch" dependence between individual actions in equilibrium in order to estimate the model. In comparison, we accommodate a flexible information structure where individuals have private information about neighborhood characteristics besides payoffs. This leads to different forms of Bayesian Nash equilibrium, and hence qualitatively different structural equations that relate the asymptotic moments to model parameters. We also allow for richer contextual effects in payoffs than these two papers. Yang and Lee (2017) analyze social interactions where the conditional expectations about group members' behaviors are heterogeneous. The individuals have asymmetric private shocks but share common knowledge about the network/group structure. Canen, Schwartz and Song (2020) applies a behavioral approach to model games on networks where agents have partial observation of neighbor types. In comparison, the individual beliefs in our model are formulated based on a common prior as is standard in simultaneous games with incomplete information. Auerbach (2022) estimates a regression model where one covariate is an unknown function of a latent driver of link formation in a network.

Our paper does not study the strategic formation of networks. This is a related but different topic that has been studied extensively in the literature. Examples include Badev

(2021), Boucher and Mourifié (2017), Chandrasekhar and Jackson (2021), Christakis, Fowler, Imbens and Kalyanaraman (2020), de Paula, Richard-Shubik and Tamer (2018), Hsieh, Lee and Boucher (2020), Mele (2017), Menzel (2015), Miyauchi (2016) and Sheng (2020). See Chandrasekhar (2016) for an extensive review.

▶ **Roadmap.** In Section 2 we present the model and establish the existence and uniqueness of pure-strategy Bayesian Nash equilibrium. In Section 3 we present our identification strategy as the number of individuals approaches infinity while the number of observed networks is fixed and small. In Section 4, we present a two-step consistent estimator. In Section 5, we provide Monte Carlo simulation results. In Section 6 we conclude. Proofs are collected in Appendix A.

2 The Model

Denote the finite set of individuals on a network by N, and let $n \equiv \#N$ denote its cardinality. The network structure is summarized by the n-by-n matrix $g \equiv (g_{ij})_{ij \in N}$. For any $i, j \in N$, let $g_{ij} \equiv 1$ if i's payoff is affected by j's action, and $g_{ij} \equiv 0$ otherwise. By convention in the literature, let $g_{ii} \equiv 0$. Define the set of neighbors for i by $N_i \equiv \{j \in N : g_{ij} = 1\}$. Each individual has a vector characteristics x_i, which has a discrete finite support X with $\#X \equiv K$. Let $x_{N_i} \equiv (x_j)_{j \in N_i}$ denote the characteristics of i's neighbors.

Let $\tau_i \equiv (x_i, N_i, x_{N_i}, \varepsilon_i)$ summarize the information available to individual i, where $\varepsilon_i \in \mathbb{R}$ is an idiosyncratic shock to i's payoff. Each individual $i \in N$ chooses a continuous action a_i from $A = \mathbb{R}$; this rules out settings with discrete actions such as those in Lin and Xu (2017) and Xu (2018). The payoff for each individual i from choosing a_i is:

$$u_i(a_i, a_{-i}, \tau_i) \equiv \tilde{h}_i(\tau_i)a_i - \frac{1}{2}a_i^2 - \frac{\phi}{2}\sum\nolimits_{j \in N_i} \tilde{w}_{ij}(\tau_i)(a_i - a_j)^2, \tag{1}$$

where $\phi > 0$ and $a_{-i} \equiv (a_j)_{j \in N \setminus \{i\}}$; and $\tilde{w}_{ij}(\tau_i)$ are positive, bounded peer effect weights that i assigns to its own deviation from a neighbor j's choice.[3] The positive sign of ϕ means that the individuals incur utility losses due to non-conformity with the actions of their neighbors. That is, the last quadratic term in u_i indicates the costs for deviation from group peers. The

[3] If $\phi = 0$, then the model is reduced to n single-agent decisions, with the individuals making independent decisions with no strategic interaction.

function $\tilde{h}_i$ contributes to a "contextual effect" and the last term leads to a "peer effect"in the terminology of Manski (1993). The second term reflects the direct costs of the action.

We maintain the following assumption about the information available to each individual.

Assumption 1 *(Information) For each $i \in N$, $(N_i, x_{N_i}, \varepsilon_i)$ is privately known and unobserved by other individuals $j \neq i$.*

The common prior for $(\tau_i)_{i \in N}$ is given by a distribution function F, which is known to all individuals. Let T denote the support of each τ_i.[4] A pure strategy for individual i is a mapping from T to A. A pure-strategy Bayesian Nash equilibrium (p.s.BNE) is a profile of pure strategies $(s_i)_{i \in N}$ such that:

$$s_i(\tau_i) \in \arg\max_{a_i \in A} \mathbf{E}[u_i(a_i, \mathbf{s}_{-i}(\tau_{-i}), \tau_i) \mid \tau_i] \quad \forall i \in N,$$

where $\mathbf{s}_{-i}(\tau_{-i}) = (s_j(\tau_j))_{j \in N \setminus \{i\}}$. The expectation $\mathbf{E}$ integrates out $\tau_{-i} \equiv (\tau_j)_{j \in N \setminus \{i\}}$ with respect to its conditional distribution given τ_i, as implied by the common prior F. Assuming the order of integration and differentiation can be swapped, we use the first-order condition for each i to derive the following best response to $\mathbf{s}_{-i} \equiv (s_j)_{j \in N \setminus \{i\}}$:

$$R_i(\tau_i; \mathbf{s}_{-i}) = \frac{\tilde{h}_i(\tau_i) + \phi \sum_{j \in N_i} \tilde{w}_{ij}(\tau_i) \mathbf{E}\left[s_j(\tau_j) \middle| \tau_i \right]}{1 + \phi \sum_{j \in N_i} \tilde{w}_{ij}(\tau_i)}.$$

A p.s.BNE can be characterized through the following fixed-point equation:

$$s_i(\tau_i) = R_i(\tau_i; \mathbf{s}_{-i}) \text{ for all } i \in N \text{ and } \tau_i.$$

An argument similar to Blume et al (2015) implies there exists a unique p.s.BNE in this network game.

Theorem 1 *(Uniqueness of p.s.BNE) Under Assumption 1, there exists a unique p.s.BNE.*

We prove the theorem by showing that the best response mapping is a contraction. The positive sign of ϕ means the individuals value conformity with peers. If $\phi = 0$, then payoffs in (1) would consist of no peer effects, thus reducing the model to *single-agent* choices.[5]

[4] Theorem 1 continues to hold when the support of τ_i differs across $i \in N$.

[5] Note the existence and uniqueness of equilibrium is immediate in this case, because each agent is maximizing a quadratic utility independent of each other with no interactions.

The rest of this section focuses on an anonymous version of the game above, where only the profile of characteristics (rather than actual identities) of neighboring individuals affect an individual's ex post payoffs and interim beliefs. We show that in such a context, the unique p.s.BNE is symmetric (i.e., all individuals share the same pure strategy) and anonymous (i.e., the equilibrium strategy only depends on the profile of characteristics but not the actual identities). For each $x \in X$, let $N_{i,x} \equiv \{j \in N_i : x_j = x\}$ denote the set of neighbors of i whose characteristic is x. Let $n_{i,x} \equiv \# N_{i,x}$ for each $x \in X$; and let a K-vector $\mathbf{n}_i \equiv (n_{i,x})_{x \in X}$ summarize the distribution of x_j in the neighborhood N_i. By construction, $\sum_{x \in X} n_{i,x} = \# N_i$.

We consider the class of games where individuals' payoffs and interim beliefs depend on some sufficient statistics of $\mathbf{n}_i$. Specifically, let $\pi : \mathbb{N}^K \to M$ be a function that summarizes neighbor characteristics, with its range M being in the Euclidean space and encompassing possible realized values under all sample sizes. For example, $\pi(\mathbf{n}_i) \in M$ may be some moment based on the empirical distribution of $(x_j)_{j \in N_i}$, or may report *censored* frequencies of neighbors with certain types. In our simulations in Section 5, we investigate two data-generating processes with different forms of $\pi(\cdot)$. In one case, $\pi(\mathbf{n}_i)$ returns the number of neighbors with each type, censored at certain caps; in the other, it discretizes the average deviation from neighbors' characteristics. It is important to note that $\pi(.)$ is a function of $\mathbf{n}_i$ but not of the specific identities of neighbors in N_i, and that this function does not vary with sample size n. In what follows we use $m_i \equiv \pi(\mathbf{n}_i)$ as shorthand. In practice we choose to define $\pi(\mathbf{n}_i)$ as a lower-dimensional summary statistic of $\mathbf{n}_i$ for tractability in estimation.

Assumption 2 *(Symmetry and Sufficiency in Payoffs) (i) There exists $h : X \times M \times \mathbb{R} \to \mathbb{R}$ such that $\tilde{h}_i(\tau_i) = h(x_i, m_i, \varepsilon_i)$ for all i and τ_i, and $\mathbf{E}[h(x_i, m_i, \varepsilon_i) | x_i = x, m_i = m]$ exists for all $x \in X$, $m \in M$. (ii) For each i, $\tilde{w}_{ij}(\tau_i) = w(x_j, x_i, m_i)/n_{i,x_j}$ if $j \in N_i$, and $\tilde{w}_{ij}(\tau_i) = 0$ if $j \notin N_i$, where $w(\cdot)$ is a positive function with $\sum_{x \in X} w(x, x_i, m_i) = 1$ for all (x_i, m_i).*

Assumption 2 includes direct restrictions on each individual's payoffs as well as their information set. It implies that individual payoffs and peer weights only depend on the profile of neighbor characteristics, but not on neighbor identities. This condition leads to a tractable structural model by practically reducing the complexity of each individual's payoff function, and helps to simplify the definition and use of asymptotic moments below. Later, for point identification of model primitives (Proposition 3 and 4), we further reduce the dimension of arguments here through an index sufficiency in Assumption 8.

Condition (i) states that the contextual effect for an individual i is determined by its own characteristics x_i, its private shocks ε_i, and the profile of neighbor characteristics $(x_j)_{j \in N_i}$. Thus the specific identities of individuals in N_i do not matter for contextual effects.

Condition (ii) consists of two restrictions on the peer effect weights. First, each individual i assigns equal weights to neighbors with the same characteristics x_j. Second, these weights are determined by the individual's own characteristic x_i, the neighbor's characteristic x_j, and the summary of neighbor characteristics in N_i. Together, they imply the weights do not depend on specific identities of the neighbors.

Under Assumption 2, the peer effect (third term) in (1) can be written as:

$$-\frac{\phi}{2} \sum\nolimits_{\{x: N_{i,x} \neq \emptyset\}} w(x, x_i, m_i) \Delta_i(x),$$

where $\Delta_i(x)$ denotes the mean squared difference between i's action and those of neighbors with $x_j = x$. That is, $\Delta_i(x) \equiv \frac{1}{n_{i,x}} \sum_{j \in N_{i,x}} (a_i - a_j)^2$. It is worth noting that by construction the form of the weight function $w_i(\cdot)$ does not change with the sample size n, because of the range of $\pi(\cdot)$ that defines m_i. In what follows, we define as an individual i's *anonymized* information as

$$t_i \equiv (x_i, m_i, \varepsilon_i), \text{ where } m_i \equiv \pi(\mathbf{n}_i).$$

In addition, we maintain an anonymity condition on the interim beliefs.

Assumption 3 *(Anonymity in Common Prior) The common prior F is exchangeable in the identities of individuals $i \in N$.*[6] *Under this prior, the conditional distribution of t_j given τ_i and $j \in N_i$ depends on τ_i only through (x_j, x_i, m_i).*

This condition restricts parts of the information available to the individuals (i.e., their initial priors). It lends itself to a tractable structural model by practically reducing the complexity of information processed by each individual.[7]

[6] A distribution of a random vector $(Y_1, Y_2, ..., Y_n)$ is *exchangeable* if its joint distribution is the same as that of $(Y_{\rho(1)}, Y_{\rho(2)}, ..., Y_{\rho(n)})$ for any permutation $\rho(.)$.

[7] We consider a single, large-network setting in the sample. In this case, the exact labels of individual units are not relevant for the analysis, as long as each unit is assigned a unique label.

Under Assumption 3, (x_j, x_i, m_i) are sufficient statistics for i's interim belief of a neighbor j's anonymized information t_j. That is, while individuals may not know the full identities of their second-order, indirect connections, they are likely to hold rational beliefs about the characteristics of these indirect connections. Intuitively, such beliefs only relate to the characteristics, but not identities, of the immediate neighbors. This demonstrates a similar anonymity to that in $h(\cdot)$ under Assumption 2.

With a slight abuse of notation, we write i's interim belief about t_j conditional on τ_i and $j \in N_i$ compactly as $F(t_j|x_j, x_i, m_i, j \in N_i)$ below so as to simplify notation. Exchangeability of the prior also implies this interim belief is the same across all individuals, and thus does not need to be indexed by i. Menzel (2015) used a similar exchangeability condition on the individuals' characteristics and private signals to estimate large Bayesian games. In our context, an individual's private information consists of both the payoff shocks and the neighborhood profile m_i.

In a *symmetric p.s.BNE* all individuals adopt the same pure strategy s, which maps from an individual's anonymized information t_i to A and solves the following fixed-point equation:

$$s(t_i) = r\{s\}(t_i) \text{ for all } t_i, \tag{2}$$

where r is the best response function implied by the first-order condition, and $r\{s\}$ denotes an operator on the space of symmetric pure strategies defined by:

$$r\{s\}(t_i) = \gamma h(t_i) + (1-\gamma)\sum\nolimits_{\{x:N_{i,x}\neq\emptyset\}} w(x, x_i, m_i)\mathbf{E}\left[s(t_j)|x_j = x, x_i, m_i, j \in N_i\right], \tag{3}$$

where $\gamma \equiv 1/(1+\phi)$, and the peer (interaction) effects are weighted sums over neighbor characteristics as represented in m_i. Note that, for simplicity, we use the same notation $s(\cdot)$ to denote symmetric pure strategies in (3), which only depend on an individual's anonymized information.

Corollary 1 *(Unique symmetric p.s.BNE) Suppose Assumptions 1, 2 and 3 hold. Then the unique p.s.BNE is symmetric, and the equilibrium strategies depend only on the anonymized information.*

The proof of this corollary is similar to that of Theorem 1 and is omitted for brevity. The main idea is to show that r is a contraction mapping, and then apply the Banach Fixed-point Theorem to show that $s = r\{s\}$ has a unique solution.

3 Identification

We now discuss the identification of parameters (γ, h, w) as the number of individuals approaches infinity while the number of observed networks is fixed and small. Such a dimension of asymptotics differs qualitatively from standard cross-sectional econometric models.

Suppose a researcher collects data from a *single* network with $\#N = n$ individuals. The data reports the choice a_i, the characteristics x_i and the neighbor information (N_i, x_{N_i}) for each single individual $i \in N$. Throughout this section, we maintain that such a data set with $\#N = n$ individuals is a single, random draw from some data-generating process (DGP) indexed by n. For each n, the DGP is summarized by a joint distribution F_n of the n-tuple $(\tau_i)_{i \in N_n}$. This is the common prior and satisfies the exchangeability and anonymity conditions in Assumption 3.

We establish the identification of the model via two steps. In the first step, we show that the sample average of choices made by n observed individuals on the network converge in probability to certain asymptotic moments as $n \to \infty$. We then argue that the model can not be identified from these moments without any parametric or shape restrictions on the contextual effect, even when weights in the peer effects are known to the researcher. In the second step, we show that under a mild restriction of index sufficiency on the individual weights in peer effects the model parameters can be uniquely recovered from these asymptotic moments. The identification strategy is constructive, and leads to a two-step M-estimator. In this sense, our approach is an example of "extremum-based" identification, as defined in Lewbel (2019).

3.1 Asymptotic moments

Let N_n denote a sequence of sets such that $N_n \subset N_{n'}$ for all $n < n'$. For each n, let $\mathbf{E}_n(.)$ denote the expectation under F_n. First, we define the asymptotic moments to be used in our identification analysis. Throughout this section we use $\sum_i$ as shorthand for $\sum_{i \in N_n}$, which is a sum over the n individuals in N_n; and we use $\sum_{j \neq i}$ as shorthand for $\sum_{i \in N_n} \sum_{j \in N_n \setminus \{i\}}$, which is a sum over $n(n-1)$ ordered pairs from N_n. We maintain the following condition on idiosyncratic shocks in individual payoffs.

Assumption 4 *(Exogeneity) For all n and $i, j \in N_n$, the common prior F_n is such that the payoff shock ε_j is independent from $(\varepsilon_i, x_i, m_i, g_{ij})$ conditional on (x_j, m_j).*

This assumption states that conditioning on an individual j's characteristics x_j and neighborhood profile m_j purges any correlation between its payoff shock ε_j and neighbors' characteristics or links. This exogeneity assumption requires that ε_j be conditionally independent from any idiosyncratic noises that affect link formation. It fails, for example, if the network formation process depends on unobservable variables correlated with ε_i. We consider the case where the set of neighbor profiles M is discretized and finite.

Assumption 5 *(Existence of Limits) For any $x, x' \in X$, $m, m' \in M$,*

$$\begin{aligned}
h^*(x,m) &\equiv \lim_{n\to\infty} \tfrac{1}{n} \sum_i \mathbf{E}_n[h(t_i)|x_i = x, m_i = m], \\
p^*(x,m) &\equiv \lim_{n\to\infty} \tfrac{1}{n} \sum_i \mathbf{E}_n\left[1\{x_i = x, m_i = m\}\right], \\
q^*(m'|x',x,m) &\equiv \lim_{n\to\infty} \tfrac{1}{n(n-1)} \sum_{j\neq i} \mathbf{E}_n(1\{m_j = m'\}|x_j = x', x_i = x, m_i = m, g_{ij} = 1)
\end{aligned}$$

exist and $p^*(x,m) \neq 0$.

Existence of the limit $h^*(x,m)$ is a mild restriction on the conditional distribution of ε_i. It holds, for example, if $h(t_i)$ is additively separable in ε_i and $\mathbf{E}_n(\varepsilon_i|x_i = x, m_i = m) = 0$ for all n. In Appendix B we provide an example of a random link formation process in which p^*, q^* exist.

The next proposition relates the parameters γ, w, h to asymptotic moments λ^* and q^*.

Proposition 1 *Suppose Assumptions 1, 2, 3, 4 and 5 hold. Then*

$$\lambda^*(x,m) \equiv \lim_{n\to\infty} \tfrac{1}{n} \sum_i \mathbf{E}_n(a_i|x_i = x, m_i = m) \tag{4}$$

exists and

$$\lambda^*(x,m) = \gamma h^*(x,m) + (1-\gamma) \sum_{x'\in X} w(x',x,m) \left[\sum_{m'\in M} \lambda^*(x',m') q^*(m'|x',x,m)\right] \tag{5}$$

for all $x \in X$, $m \in M$.

Equation (5) is an empirical analog of the moments implied by p.s.BNE in (2), with individuals' interim expectation about others' actions replaced by an expression that only consists of estimable asymptotic moments. The double sum on the right-hand side of (5) is an expectation of λ^* with respect to a joint distribution over a neighbor's characteristics x', m', defined as $w(., x, m) \times q^*(.|., x, m)$. Such a distribution is an individual's *weighted interim belief in the limit*, because it applies the weights in peer effects to the limit of individual interim beliefs about neighbors' m_j.

Next, we show that the asymptotic moments λ^* and q^* in (4) can be consistently estimated using sample averages across individuals *from a single network* as $n \to \infty$ under certain condition of asymptotic uncorrelation. For any $x \in X$ and $m \in M$, let $\iota_i(x, m)$ be shorthand for $1\{x_i = x, m_i = m\}$. Let $\mathbf{C}_n$ and $\mathbf{V}_n$ denote covariance and variance under F_n respectively.

Assumption 6 *(Asymptotic Uncorrelation) For any $x, x' \in X$ and $m, m' \in M$,*

(i) $\mathbf{C}_n(\iota_i(x, m), \iota_j(x, m)) \to 0$ for all $i \neq j$ as $n \to \infty$;

(ii) $\mathbf{C}_n(\iota_i(x, m)\iota_j(x', m')g_{ij}, \iota_k(x, m)\iota_\ell(x', m')g_{k\ell}) \to 0$ as $n \to \infty$ if $\{i, j\} \cap \{k, \ell\} = \emptyset$;

(iii) $\mathbf{V}_n [a_i\iota_i(x, m)]$ and $\mathbf{V}_n[(\iota_i(x, m)\iota_j(x', m')g_{ij}]$ exist for all n, and are both $o(n)$.

This assumption consists of several restrictions on the unconditional correlation between arbitrary pairs of individuals, treating their characteristics and the network links as random draws from a triangular array that is indexed by network size n. It requires that as the network size increases, the correlation between the neighborhood profiles m_i and m_j for any two individuals i and j diminishes, even though in a fixed sample two individuals i and j may share common neighbors. Also note that these restrictions are *not* conditioning on i and j being neighbors, and thus do not immediately rule out any pattern of homophily. These conditions are justified in networks where all links are directed and formed independently. Besides, in Appendix B we provide an example of a simple random Poisson network with *undirected* links in which these restrictions hold under primitive conditions such as independent, dyadic links form at diminishing rates while the expected number of neighbors for each individual converges to constants as the network size grows. This rules out certain types of dense networks. In general, these conditions also rule out scale-free networks as defined in Caldarelli, Capocci, De Los Rios, and Muñoz (2002). More broadly speaking, this condition may not hold if the neighbor profiles (m_i, m_j) for linked individuals, (i, j) with $g_{ij} = 1$, have persist correlation even as the network grows (as is the case with dyadic link formation with constant rates).

The next proposition shows that λ^* and q^* can be estimated consistently as the number of individuals on the single network in data approaches infinity ($n \to \infty$).

Proposition 2 *Suppose Assumptions 1, 2, 3, 4, 5 and 6 hold. Then*

$$\frac{\sum_i a_i \iota_i(x,m)}{\sum_i \iota_i(x,m)} \xrightarrow{p} \lambda^*(x,m)$$

and

$$\frac{\sum_{j\neq i} \iota_j(x',m')\iota_i(x,m)g_{ij}}{\sum_{j\neq i} 1\{x_j = x'\}\iota_i(x,m)g_{ij}} \xrightarrow{p} q^*(m'|x',m,x)$$

as $n \to \infty$ *for any* $x, x' \in X, m, m' \in M$ *on the support of* p^*.

A generic vector of parameters (γ', h', w') is *observationally equivalent* to the actual parameter (γ, h, w) in the DGP *based on the asymptotic moments* λ^* and q^* if (γ', h', w') satisfies (5) almost surely $[p^*]$, where λ^*, q^* are identified as probability limits in Proposition 2. We say (γ, h, w) is *identified* based on these asymptotic moments if there exists no *other* element (γ', h', w') in the parameter space that is observationally equivalent to (γ, h, w).

It is clear from (5) that (γ, h, w) can not be identified using these asymptotic moments λ^* and q^* without further restrictions. To see this, note that for any generic weight function w' (not necessarily equal to the actual weight function w) there always exist γ' and h' such that (5) holds almost surely $[p^*]$, with λ^*, q^* fixed and identified from the data. Similarly, for other $(\gamma', h') \neq (\gamma, h)$, one may construct a weight function w' that satisfies (5) by redistributing weights across realized values of x' conditional on each (x, m).

3.2 Preview of identification strategy

In this subsection we illustrate the main idea for identification using a simple example. For the rest of Section 3, we maintain Assumptions 1-6, so that the asymptotic moments are consistently estimable and are considered known for identification purpose. With this established, our strategy for model identification is similar to the case with a large number of independent static Bayesian games between a fixed number of players such as Bajari, Hong, Krainer and Nekipelov (2010). That is, we also exploit some exclusion restrictions and rank conditions in the structural form. Our conditions are invoked on peer effect weights, and therefore specific to the context of network games.

The non-identification result mentioned above shows that the parameters can not be recovered from asymptotic moments without further restrictions on the contextual effects h *and* the weights in peer effects w. Thus we focus on a semiparametric model with conditional mean restriction on the contextual effects.

Assumption 7 *(Mean Contextual Effects) $h(t_i) = \eta(x_i, m_i; \theta) + \varepsilon_i$ where η is known up to a finite-dimension parameter θ and $\mathbf{E}_n(\varepsilon_i | x_i, m_i) = 0$ for all n and $x_i \in X$, $m_i \in M$.*

This assumption states that the conditional mean of contextual effects given (x_i, m_i) is known up to a finite-dimensional parameter θ. It is commonly used in econometric models estimated by Non-linear Least Squares (NLS) or Generalized Method of Momments (GMM). Under this condition, $h^*(x, m) = \eta(x, m; \theta)$. We show the model is identified under some exclusion restriction on the weights in the peer effects as well as some rank condition on the support of observables.

▶ **Exclusion restriction.** Suppose $x \equiv (z, v)$ and has a discrete support $X = Z \times V$ where $Z \equiv \{z_{(1)}, z_{(2)}\}$ and $V \equiv \{v_{(1)}, .., v_{(\kappa)}\}$. By construction $\#X = 2\kappa \geq 4$. Also suppose that $w(x', x, m)$ is a function of z and z' alone.

For $k, \ell = 1, 2$, let $\omega_{k\ell}$ denote the weights for $z = z_{(k)}$ and $z' = z_{(\ell)}$, and let $\lambda_k(v, m)$ and $\eta_k(v, m; \theta)$ be shorthand for $\lambda^*(z_{(k)}, v, m)$ and $\eta(z_{(k)}, v, m; \theta)$. Note the weights ω_{kl} do not depend on v because of the exclusion restriction introduced in the preceding paragraph. Under this exclusion restriction, the structural link in Proposition 1 is reduced to

$$\lambda_k^*(v, m) = \gamma \eta_k(v, m; \theta) + (1 - \gamma) \sum\nolimits_{\ell=1,2} \omega_{k\ell} \sum\nolimits_{v', m'} Q_{k\ell}^*(v', m', v, m) \lambda_\ell^*(v', m') \tag{6}$$

for $k = 1, 2$, where the summation $\sum_{v', m'}$ is over the supports V and M, and

$$Q_{k\ell}^*(v', m', v, m) \equiv \frac{q^*(m' \mid z' = z_{(\ell)}, v', z = z_{(k)}, v, m)}{\#V} \text{ for } \ell, k = 1, 2.$$

That is, $Q_{k\ell}^*$ summarizes an individual i's belief about a neighbor j's own neighborhood profile m_j, based on i's information (x_j, x_i, m_i) and adjusted by the weights that i assigns in peer effects. By construction, $\omega_{k\ell} Q_{k\ell}^*(., ., v, m)$ is an individual i's weighted interim belief about v', m' and $z' = z_{(\ell)}$ conditional on $(z_i, v_i, m_i) = (z_{(k)}, v, m)$, as explained after Proposition 1.

▶ **Rank condition.** To fix ideas, suppose (x, m) are discrete and denote $\lambda_k^*(.)$, $\eta_k(.,.;\theta)$ and $Q_{k\ell}^*(.,.)$ respectively by two column vectors λ_k^*, $\eta_k(\theta)$ and a square matrix $Q_{k\ell}^*$, with each component in λ_k^* and $\eta_k(\theta)$ and each row and column in $Q_{k\ell}^*$ corresponding to an element (v, m) on the joint support $V \times M$. Let $\lambda^* \equiv [\lambda_1^{*\prime}, \lambda_2^{*\prime}]'$, $\eta(\theta) \equiv [\eta_1(\theta)', \eta_2(\theta)']'$. Then we can write (6) as:

$$\lambda^* = \gamma\eta(\theta) + (1-\gamma)Q_\omega^*\lambda^*, \tag{7}$$

where

$$Q_\omega^* \equiv \begin{pmatrix} \omega_{11}Q_{11}^*, & \omega_{12}Q_{12}^* \\ \omega_{21}Q_{21}^*, & \omega_{22}Q_{22}^* \end{pmatrix}.$$

Equation (7) consists of $2 \times \#V \times \#M$ equalities and involves unknown parameters γ, θ, $\omega \equiv (\omega_{k\ell})_{k,\ell=1,2}$. as well as the identified asymptotic moments $\lambda_k^*(.)$ and $q^*(.|.)$ (in $Q_{k\ell}^*$). These equalites are "quasi-structural" in that they depend on the expected choices of actions λ_k^*, which themselves are endogenous objects arising from the equilibrium. However, by Proposition 2, both λ^*, q^* are consistently estimable from sample averages and can be considered known in identification.[8]

We now derive the rank conditions needed for uniquely recovering γ, θ, ω from (7). Suppose there exists some other $(\gamma', \theta', \omega') \neq (\gamma, \theta, \omega)$ that is observationally equivalent to (γ, θ, ω). Then the right-hand side of (7) must remain the same when (γ, θ, ω) is replaced by $(\gamma', \theta', \omega')$ based on the asymptotic moments λ^* and q^*. This implies there exists a column vector $v_k \in \mathbb{R}^4 \backslash \{0\}$ such that $[\eta_k(\theta'), \eta_k(\theta), Q_{k1}^*\lambda_1^*, Q_{k2}^*\lambda_2^*]v_k = 0$ for $k = 1, 2$. An intuitive condition that prevent this from happening is:

$$\text{"For any } \theta' \neq \theta, [\eta_k(\theta'), \eta_k(\theta), Q_{k1}^*\lambda_1^*, Q_{k2}^*\lambda_2^*] \text{ has full rank for } k = 1, 2\text{"}. \tag{8}$$

Thus this is a sufficient condition for identifying (γ, θ, ω) from the asymptotic moments. In general it holds when the conditional mean contextual effects η is nonlinear in θ and (x, m), and there is enough variation of (x, m) on the support. Also note that because the functional form of η_k is known (up to θ) and $\lambda_k^*, Q_{k\ell}^*$ are identified from Proposition 2, the rank condition in (8) can be tested.

[8]The econometrics literature abounds in examples where structural models are identified using quasi-structural equations which involve equilibrium outcomes. For instance, see Bajari, Hong, Krainer and Nekipelov (2009) in the context of static discrete games with incomplete information; and Aguirregabiria and Mira (2010) in dynamic games with Markovian perfect equilibria.

▶ **Special case: no peer effects ($\phi = 0$).** So far, we have assumed $\phi > 0$. If instead $\phi = 0$, then the model would be reduced to single-agent decisions, and we would not be able to identify the weights ω_{kl}. Formally, if $\phi = 0$, then the second term on the right-hand side of (7) would disappear and we would not be able to recover ω_{kl} under the rank conditions.

If a researcher does not know whether ϕ is strictly positive or zero, then a null hypothesis $\phi = 0$ can be tested by regressing individual actions a_i on the sample analog of peer effects in a model with some parametrization, such as that introduced later in (5), and then checking the statistical significance of the coefficient in front of the latter.[9]

▶ **Special case: linearity in parameters.** The rank condition in (8) does not hold when η is linear in θ. That is, $\eta_k(\theta) = \zeta_k\theta_k$ for all $k = 1, 2$, where ζ_k is a $(\#V \times \#M)$-by-dim(θ_k) matrix of known functions of (v, m), and $\theta_k \neq 0$ is a vector of constant coefficients. In such cases, $[\zeta_k\theta'_k, \zeta_k\theta_k]$ can not have full rank for any θ'_k that is proportional to θ_k. Nevertheless it is relatively straight-forward to adjust the argument above to derive the following rank conditions for identifying (γ, θ, ω):

$$\text{“}[\zeta_k, Q^*_{k1}\lambda^*_1, Q^*_{k2}\lambda^*_2] \text{ has full rank for } k = 1, 2\text{”}. \tag{9}$$

To see why, suppose there exists some $(\gamma', \theta', \omega')$ that is observationally equivalent to (γ, θ, ω). Then there exists a column vector $\tau_k \in \mathbb{R}^3\backslash\{0\}$ such that $[\zeta_k, Q^*_{k1}\lambda^*_1, Q^*_{k2}\lambda^*_2]\tau_k = 0$ for $k = 1, 2$. Thus the rank condition (9) is sufficient for identification.

It follows from (7) that the reduced-form for asymptotic moments in equilibrium is:

$$\lambda^* = \gamma[I - (1-\gamma)Q^*_\omega]^{-1} \begin{pmatrix} \zeta_1\theta_1 \\ \zeta_2\theta_2 \end{pmatrix}.$$

Hence the rank condition in (9) can be expressed in terms of model primitives:

$$\text{“}[\zeta_k,\ \gamma Q^*_{k1}(M_{11}\zeta_1\theta_1 + M_{12}\zeta_2\theta_2),\ \gamma Q^*_{k2}(M_{21}\zeta_1\theta_1 + M_{22}\zeta_2\theta_2)] \text{ has full rank for } k = 1, 2\text{”}, \tag{10}$$

where $M_{k\ell}$ for $k, \ell = 1, 2$ are conformable submatrices partitioning the inverse of $I - (1-\gamma)Q^*_\omega$.

▶ **A Numerical Example.** We conclude this preview with a numerical example that illustrates the rank conditions. Let $x_i \equiv (z_i, v_i)$, where v_i and z_i are binary with supports

[9] Large-sample property of such test statistics would need to account for first-stage estimation errors in the peer effects, as is typical in two-step M-estimation. We leave this as a direction for future research.

$\{1, 2\}$ and $\{z_{(1)}, z_{(2)}\}$ respectively. Define $m_i \equiv 1\left\{\frac{\#\{j \in N_i : z_j = z_{(1)}\}}{\#N_i} \geq \frac{1}{2}\right\}$. That is, the contextual effect depends on the network structure only through the proportion of neighbors with $z_j = z_{(1)}$. Suppose $\eta(x_i, m_i; \theta) \equiv \theta_d v_i m_i$ if $z_i = z_{(d)}$ for $d = 1, 2$, and suppose that the weights in peer effects only depend on z_j, z_i alone. Let $\theta_1 = 0.8$, $\theta_2 = 1.3$, $\gamma = 0.7$, $\omega_{11} = 0.6$, $\omega_{12} = 0.4$, $\omega_{21} = 0.3$, $\omega_{22} = 0.7$. Let

$$Q^*_{k\ell} = \begin{pmatrix} 0.15 & 0.40 & 0.25 & 0.20 \\ 0.15 & 0.40 & 0.25 & 0.20 \\ 0.40 & 0.10 & 0.30 & 0.20 \\ 0.40 & 0.10 & 0.30 & 0.20 \end{pmatrix} \forall k, \ell = 1, 2,$$

where the (i, j)-entry in $Q^*_{k\ell}$ corresponds to $Q^*_{k\ell}(v', m', v, m)$ with (v, m), (v', m') being the i-th and j-th element in $\{(1, 1), (1, 0), (2, 1), (2, 0)\}$. It is worth emphasizing that in our specification of $Q^*_{k\ell}$ above, we intentionally minimize the source of exogenous variation by restricting $q^*(m' \mid z', v', z, v, m)$ to be invariant in (z', z, m). Yet even in this scenario it is straightforward to verify that the rank condition in (10) holds.

3.3 Formal results: index sufficiency

In this subsection we generalize and formalize the identification argument in Section 3.2. Our method requires an index sufficiency condition on the weights in peer effects.

Assumption 8 *(Index Sufficiency) There exist known indexes* $\psi : X \to \Psi$ *and* $\varphi : X \times M \to \Phi$*, where* $\dim(\Psi) < \dim(X)$ *and* $\dim(\Phi) \leq \dim(X \times M)$*, such that* $w(x', x, m) = w(y', y, \tilde{m})$ *whenever* $\psi(x') = \psi(y')$ *and* $\varphi(x, m) = \varphi(y, \tilde{m})$ *for all* $x, x', y, y' \in X$ *and* $m, \tilde{m} \in M$*.*

Index sufficiency is used frequently in semiparametric econometric models. (See Powell (1994) for further discussion.) In our context, Assumption 8 is decomposed into two substantive restrictions. First, individuals with the same index $\varphi(x_i, m_i)$ assigns weights to neighbors in the same way. Second, neighbors with the same index $\psi(x_j)$ always receive the same weight. An intuitive special case of such index sufficiency is the exclusion restriction mentioned in Section 3.2 with with the indexes $\psi(x')$ and $\varphi(x, m)$ being subvectors of x' and (x, m) respectively.

Under this condition, we can reparametrize the weight function w as a function defined over the lower-dimensional support of indexes. That is, there exists $\omega : \Psi \times \Phi \to [0,1]$ with $\sum_{c\in\Psi} \omega(c,d) = 1$ for all $d \in \Phi$ such that:

$$w(x', x, m) = \frac{\omega(\psi(x'), \varphi(x,m))}{\#\{\tilde{x} : \psi(\tilde{x}) = \psi(x')\}} \text{ for all } x', x \in X \text{ and } m \in M.$$

Under Assumptions 7 and 8, the link between asymptotic moments and parameters in (5) is

$$\lambda^*(x,m) = \gamma\eta(x,m;\theta) + (1-\gamma)\sum_{c\in\Psi} \omega(c, \varphi(x,m))\mu^*(c,x,m), \tag{11}$$

where

$$\mu^*(c,x,m) \equiv \frac{\sum_{\{x':\psi(x')=c\}} \sum_{m'\in M} \lambda^*(x',m')q^*(m'|x',x,m)}{\#\{\tilde{x} : \psi(\tilde{x}) = c\}}.$$

Because the index function ψ is known, Proposition 2 implies that μ^* is identified (and consistently estimable) as $n \to \infty$.[10]

By definition, the true model elements (γ, θ, ω) in the data-generating process is identified if the equality in (11) fails at least for some set of (x, m) with positive measure in p^* whenever (γ, θ, ω) is replaced by a different vector of parameters $(\tilde{\gamma}, \tilde{\theta}, \tilde{\omega}) \neq (\gamma, \theta, \omega)$.

To fix ideas, suppose $\#M < \infty$ so that $\#\Psi < \infty$ and $\#\Phi < \infty$. For each (x, m), let $\mu^*(x,m) \equiv (\mu^*(c,x,m))_{c\in\Psi}$ be a row-vector. We say a random row-vector $\mathbf{v}$ *has full rank conditional on some event* $\mathcal{E}$ (under a probability measure p) if there exists no column-vector $\alpha \neq \mathbf{0}$ such that $p\{\mathbf{v}\alpha = 0 \mid \mathcal{E}\} = 1$.

Assumption 9 *(Rank Condition) For any $\tilde{\theta} \neq \theta$ and any $d \in \Phi$, $[\eta(x,m;\tilde{\theta}), \eta(x,m;\theta), \mu^*(x,m)]$ has full rank conditional on $\varphi(x,m) = d$ under p^*.*

This condition requires there be sufficient variation over an individual's private information (x, m). Generalization to the case with $\#M{=}\infty$ would involve some form of "completeness" condition on linear operators defined by integrals.

[10]To derive (11), use the reparametrized weights ω to write $\lambda^*(x,m)$ as

$$\begin{aligned} & \gamma h(x,m) + (1-\gamma)\sum_{c\in\Psi}\sum_{\{x':\psi(x')=c\}} \tfrac{\omega(c,\varphi(x,m))}{\#\{\tilde{x}:\psi(\tilde{x})=c\}} \Big[\sum_{m'} \lambda^*(x',m')q^*(m'|x',x,m)\Big] \\ = \; & \gamma h(x,m) + (1-\gamma)\sum_{c\in\Psi} \omega(c,\varphi(x,m)) \tfrac{\sum_{\{x':\psi(x')=c\}}\sum_{m'} \lambda^*(x',m')q^*(m'|x',x,m)}{\#\{\tilde{x}:\psi(\tilde{x})=c\}}, \end{aligned}$$

where the equality holds because $\omega(c, \varphi(x,m))$ is constant over $\{x' : \psi(x') = c\}$ and $m' \in M$.

Proposition 3 *Suppose Assumptions 7, 8 and 9 hold. Then* (γ, θ, ω) *are identified from the asymptotic moments.*

Assumption 9 is often satisfied when η is nonlinear in θ as well as (x, m). However, the rank condition in Assumption 9 does not hold in a class of models that are linear in parameters. These are models with $\eta(x, m; \theta) = \zeta_d(x, m)'\theta_d$ whenever $\varphi(x, m) = d$, where the vector of functions $\zeta_d : X \times M \to \mathbb{R}^{L_d}$, $L_d < \infty$ are known up to a finite-dimensional parameter θ_d.

For example, $\zeta_d(x, m) \equiv [1, x, f(x, m)]$ for all $d \in \Phi$, where $x \in \mathbb{R}^{D_X}$ and $f(x, m) : X \times M \to \mathbb{R}^{D_f}$ is a vector of component-wise squared differences between x and the mean of neighbor characteristics. Then $\eta(x, m; \theta) = \theta_{d,0} + x\theta_{d,X} + f(x, m)\theta_{d,f}$ whenever $\varphi(x, m) = d$, with $\theta_d \equiv (\theta_{d,0}, \theta_{d,X}, \theta_{d,f})$ and $L = 1 + D_X + D_f$. For such a class of models, Assumption 9 does not hold because for any $\tilde{\theta}_d \neq \theta_d$ that is proportional to θ_d, $[\zeta_d(x, m)'\tilde{\theta}_d$, $\zeta_d(x, m)'\theta_d$, $\mu^*(x, m)]$ can not have full rank conditional on $\varphi(x, m) = d$ regardless of the value of d. Our next proposition shows that for such models (θ, γ, ω) is identified under different and yet intuitive rank conditions.

Proposition 4 *Suppose Assumptions 7 and 8 hold with* $\eta(x, m; \theta) = \zeta_d(x, m)'\theta_d$ *for all* (x, m) *such that* $\varphi(x, m) = d$*, where* $\zeta_d : X \times M \to \mathbb{R}^{L_d}$ *is known,* $\theta_d \neq 0$ *and* $L_d < \infty$ *for each* $d \in \Phi$*. Then* (γ, θ, ω) *are identified from the asymptotic moments if for each* $d \in \Phi$*,* $[\zeta_d(x, m), \mu^*(x, m)]$ *has full rank conditional on* $\varphi(x, m) = d$ *under* p^**.*

4 Two-Step M-Estimator

Consider a sample that consists of n individuals on a network. Across individuals $i = 1, 2, ..., n$, the vector of individual characteristics and shocks (errors) (x_i, ε_i) are drawn independently from a fixed distribution (which does not vary with the sample size). The data-generating process satisfies the conditions on the error distribution, the link formation, as well as the individual payoffs in Assumptions 1-7.

We propose a two-step estimator for parameters in individual payoffs. First, estimate the asymptotic moments. Then, estimate the payoff parameters by matching the implied asymptotic moments with empirical analogs from the data.[11] Throughout this section, we maintain Assumptions 1 to 7 in Sections 2 and 3.

To fix ideas, we present the estimator for the example in Section 3.2, where the observed characteristics are discrete (i.e., $X = Z \times V$ with $Z \equiv \{z_{(1)}, z_{(2)}\}$ and $V \equiv \{v_{(1)}, .., v_{(\kappa)}\}$ for $\kappa \geq 2$), and the weights in peer effects only depend on the binary characteristics z_i of an individual and its neighbors. Generalization to the cases with continuous covariates in $Z \times V$ is complex and left for future research.

To simplify notation, we reparametrize the model as

$$\beta_{k\ell} \equiv (1-\gamma)\omega_{k\ell} \text{ for } k, \ell \in \{1, 2\}.$$

In what follows, we use $\delta_0 \equiv [\gamma_0; \beta_0; \theta_0]$ to denote the true parameters in the data-generating process and let $\delta \equiv [\gamma; \beta; \theta]$ denote a generic element in the parameter space D.

Partition the set of individuals N_n into $N_{n,(k)} \equiv \{i \in N_n : z_i = z_{(k)}\}$ for $k = 1, 2$, and let $\hat{\lambda}, \hat{q}$ be the non-parametric estimators for λ^*, q^* in Proposition 2. Let $\eta_i(\theta) \equiv \eta(x_i, m_i; \theta)$, $\hat{\lambda}_i \equiv \hat{\lambda}(z_i, v_i, m_i)$, $\hat{q}_{\ell,i}(m'|v') \equiv \hat{q}(m'|z' = z_{(\ell)}, v', x_i, m_i)$ and

$$\hat{\chi}_{\ell,i} \equiv \tfrac{1}{\#V} \sum\nolimits_{v',m'} \hat{q}_{\ell,i}(m'|v')\hat{\lambda}(z_{(\ell)}, v', m').$$

Our two-step estimator is:

$$\hat{\delta}_n \equiv \arg\min_{\delta \in D} \hat{G}_n(\delta)$$

with

$$\hat{G}_n(\delta) \equiv n^{-1} \sum\nolimits_i \left[\hat{\lambda}_i - \gamma\eta_i(\theta) - \sum\nolimits_k 1\{z_i = z_{(k)}\} \sum\nolimits_\ell \hat{\chi}_{\ell,i}\beta_{k\ell}\right]^2.$$

This estimator is consistent under a set of conditions T1-T3 presented and discussed below.

T1 (Parameter Space) δ_0 is in the interior of a compact parameter space D.

Compactness of D in T1 ensures that a minimum of the probability limit of the objective function exists. That δ_0 lies in the interior of the parameter space allows for Taylor series expansion around the true parameter in the proof of consistency in Proposition 5 below.

[11]This procedure is reminiscent of an estimator that Hotz and Miller (1993) proposed for dynamic discrete choice models in that they both requires estimation of certain moments in a first-step. In our case, the second step consists of m-estimation based on our identification argument.

T2 (Identification) For any $\delta \neq \delta_0$ in D, (6) does not hold for a set of (x, m) with positive measure under p^*.

Condition T2 states that the model parameters are point identified from the structural relation between these parameters and the recoverable asymptotic moments, which is summarized in (6). In Section 3, we already offered a detailed discussion of identification, and provided sets of sufficient conditions under which T2 is satisfied. Under T2, the probability limit of the objective function in our m-estimator (introduced below) is uniquely minimized at the true parameter.

T3 (Regular conditions) (i) For all $k, l \in X$, the asymptotic moments $\lambda^*(\cdot)$ and $q^*(\cdot|\cdot)$ are uniformly bounded over their domains. (ii) The mean contextual effect $\eta(.,.;\theta)$ is continuously differentiable in θ with a bounded gradient almost surely under p^*.

In the proof of estimator consistency below, we use the boundedness on the asymptotic moments in T3 (i) to show the objective function converges in probability to its population counterpart point-wise on the parameter space. The smoothness and boundedness conditions on the contextual effects in T3 (ii) are then used for strengthening the result into a uniform convergence (in probability) of the objective function.

Proposition 5 *Suppose Assumptions 1 to 7 hold. Then* $\hat{\delta}_n \xrightarrow{p} \delta_0$ *under the conditions T1, T2, T3.*

An alternative approach for two-step estimation would be as follows. First, use (5) and iterative, forward-substitution to express λ^* as an infinite series that depends on the parameters and asymptotic moments that can be estimated from the data. Then use a minimum-distance method to estimate the parameters by matching the estimated infinite series with the estimate $\hat{\lambda}$. Compared with our current approach, this alternative estimator involves higher computation costs, and more complex conditions are needed in order to show identification and consistency using its objective function.

In Appendix C, we discuss the asymptotic distribution of our two-step m-estimator. Our goal in Appendix C is not so much to conduct formal inference as to offer a framework for understanding and deriving the estimator's asymptotic properties. A formal characterization and derivation of the limiting distribution of our estimator under primitive conditions, as well as construction of valid standard errors, would require elaborate conditions that restrict the strength of network dependence. We do not address these challenges in this paper.

Leung (2021) develops an inference procedure that is robust to general forms of weak dependence in the data. His method uses resampled test statistics and does not depend on the unknown correlation structure in the sample, with leading examples being various forms of network dependence. Leung and Moon (2023) prove a central limit theorem for network moments in a model of network formation with strategic interactions and homophilous agents as a single large network in the sample grows. They show that a modification of "exponential stabilization" conditions from the literature on geometric graphs provides a useful formulation of weak dependence, which they use to establish an abstract central limit theorem. They also derive primitive conditions for stabilization using results in branching process theory, and discuss practical inference procedures. An interesting direction for future research would be to specify primitive conditions in our context that could lead to the form of weak dependence in Leung (2021) or Leung and Moon (2023) so that their inference methods can be applied in our setting.

Generalization of this estimator to the case with continuous covariates in $Z \times V$ will be complex. It typically requires us to use kernel or series estimators for the asymptotic moments in the first step. Proof of consistency in such a case involves showing the uniform convergence of these estimators of asymptotic moments over the support types and neighbor profiles. We leave a full-fledged formal analysis for the case of continuous covariates for future research.

5 Simulation

In this section we present simulation evidence for the performance of our two-step m-estimator, using simulated samples under various designs of the DGP. The contextual effect in (3) is parametrized as

$$\gamma h(t_i) = x_i\beta_x + m_i\beta_m + \epsilon_i,$$

where $\beta_x = 3.0$, $\beta_m = 1.5$, the error ϵ_i follows a zero-mean truncated normal distribution, and the individual characteristic x_i is uniformly distributed over a discrete support. The peer effects in (3) is parametrized as $1 - \gamma = 0.8$, with weights allocated equally among different types of neighbors. We experiment with two distinctive definitions of the sufficient statistic for neighbor characteristics: (i) m_i is a discretized value of $\frac{1}{\#N_i}\sum_{j\in N_i}|x_j - x_i|$; and (ii) the number of same-type neighbors censored above at 10.

For every design and sample size n considered, we simulate $S = 200$ independent samples, each of which consists of observable characteristics x_i and choices a_i by n individuals on a *single* network. In each sample, the individual characteristics are drawn independently from a specified support X. The links between individuals are undirected, and are formed independently with some probability that depends on individual characteristics. We study two scenarios for each pair x_i, x_j: (a) the link formation probability $p_n(x_i, x_j)$ decreases as the sample size n increases, and $np_n(x_i, x_j)$ converge to a constant; and (b) the link formation probability is fixed $p(x_i, x_j)$ and invariant as $n \to \infty$.

The individual choices under the symmetric pure-strategy Bayesian Nash equilibrium are simulated using the following steps. First, use our specification of the DGP to calculate individuals' interim belief about m_j conditional on x_j, m_i, x_i and $g_{ij} = 1$. Next, plug in this belief into the fixed-point characterization of equilibrium and solve for the endogenous moment $\mathbf{E}_n(a_i|x_i, m_i)$. (See the proof of Proposition 1 in Appendix A for details.) Then, draw individual noises ϵ_i from the distribution specified and set $a_i = \mathbf{E}_n(a_i|x_i, m_i) + \epsilon_i$.

For each sample, we calculate our two-step m-estimator $\hat{\beta}_x, \hat{\beta}_m, 1 - \hat{\gamma}$. No smoothing parameter is required because of the discrete support for x_i, m_i. The tables below report the empirical bias, variance and mean-squared errors (MSE) from $S = 200$ estimates. The network sizes are set at $n = 200, 400, 800$.

Table 1. Discretized m_i; $X = \{0,1\}$; $np_n \to (10,5,10)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2,3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0164	0.0135	0.0138	0.1930	0.2098	0.2470	-0.0907	0.0314	0.0396
400	0.0106	0.0055	0.0056	0.1651	0.0953	0.1225	-0.0880	0.0189	0.0264
800	-0.0007	0.0036	0.0036	0.0369	0.0210	0.0224	-0.0432	0.0096	0.0115
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2,2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0193	0.0251	0.0255	0.1886	0.2435	0.2791	-0.0910	0.0357	0.0439
400	0.0148	0.0111	0.0113	0.1046	0.1169	0.1279	-0.0738	0.0223	0.0277
800	-0.0082	0.0061	0.0061	0.0154	0.0434	0.0436	-0.0290	0.0171	0.0179

Note: The bias, variance and mean squared errors in this table are calculated using $S =$ 200 independent samples of single networks with n individuals. Neighborhood profile m_i is defined as $(\#N_i)^{-1}\sum_{j\in N_i}|x_j - x_i|$ rounded to the nearest multiple of $\frac{1}{5}$.

Table 1 reports the results for a design where x_i is Bernoulli with equal probability and m_i is the discretization of the average neighbor characteristics $(\#N_i)^{-1}\sum_{j\in N_i}|x_j - x_i|$, defined by rounding this average to the nearest multiple of $\frac{1}{5}$. As the sample size increases, the independent link-formation probability diminishes but converges to nonzero constants. For simplicity in implementation, we set $nP_n\{g_{ij} = 1|x_i = x_j = k\} = 10$ for $k \in \{0,1\}$ and $nP_n\{g_{ij} = 1|x_i \neq x_j\} = 5$ for all n. The two panels in Table 1 show how the estimator's performance varies with the support and variance of individual noises.

The MSEs for three parameters decrease as $n \to \infty$, as the consistency in Proposition 5 implies. The MSEs are quite small even for a moderate sample size $n = 800$. The estimation error in the contextual effect of m_i appears to be greater than that of the peer effects τ and individual effect of x_i. The distribution of individual payoff noises ϵ_i affects the estimation accuracy, as the MSEs are greater for models with error terms that have higher variances.

Table 2: Censored m_i; $X = \{0,1\}$; $np_n \to (10,5,20)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0460	0.0687	0.0708	0.0190	0.0047	0.0051	-0.1498	0.0127	0.0351
400	-0.0451	0.0267	0.0287	0.0111	0.0019	0.0020	-0.1130	0.0065	0.0193
800	0.0146	0.0152	0.0154	0.0089	0.0009	0.0010	-0.0890	0.0036	0.0116
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0505	0.0762	0.0787	0.0214	0.0060	0.0064	-0.1534	0.0162	0.0397
400	-0.0375	0.0346	0.0360	0.0147	0.0017	0.0019	-0.1172	0.0075	0.0212
800	0.0047	0.0176	0.0176	0.0109	0.0009	0.0011	-0.0898	0.0041	0.0122

Note: The bias, variance and mean squared errors in this table are calculated using $S = 200$ independent samples of single networks with n individuals. For each i, the neighbor profile m_i is defined as the minimum of same-type neighbors and 10.

Table 2 reports the results for a similar model where the neighbor profile is defined as the censored number of same-type neighbors. As before, the MSEs converge to zero as $n \to 0$. In contrast, the estimator for the contextual effect of m_i is more accurate than the case where m_i is the discretization of average neighbor characteristics. This may be explained in part by a richer variation in m_i under this new specification with $np_n \to (10, 5, 20)$ relative to that in Table 1. While we do not provide a formal result about the rate of convergence of our two-step estimator, the rate of convergence in the MSE in both Table 1 and Table 2 appears to be reasonably close to $\sqrt{n}$ with few exceptions in the finite sample.

Table 3: Censored m_i; $X = \{0, 1\}$; $np_n \to (20, 10, 20)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0349	0.0261	0.0273	0.0094	0.0020	0.0020	-0.0708	0.0033	0.0083
400	0.0255	0.0158	0.0165	0.0054	0.0011	0.0011	-0.0662	0.0021	0.0065
800	-0.0193	0.0093	0.0097	-0.0037	0.0004	0.0004	-0.0459	0.0010	0.0031
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0446	0.0367	0.0387	0.0116	0.0037	0.0038	-0.0758	0.0062	0.0119
400	0.0313	0.0214	0.0224	0.0076	0.0015	0.0016	-0.0708	0.0037	0.0087
800	-0.0248	0.0113	0.0119	-0.0037	0.0007	0.0007	-0.0450	0.0015	0.0035

Table 3 reports results under a design almost identical to that in Table 2, except that the sequence of link-formation probability now converges to a higher level $(20, 10, 20)$. In comparison with Table 2, the MSE in this case are slightly smaller. This pattern is related to the fact that a higher link formation probability tends to increase the variation in neighbor profile defined as the censored number of same-type neighbors. Similar to Tables 1 and 2, the results in Table 3 demonstrates that an increase in the variance of noises leads to slightly worse performance of the estimators.

Table 4: Censored m_i; $X = \{0, 1, 2\}$; $np_n \to (10, 8, 5, 12, 8, 10)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0368	0.0228	0.0241	0.0587	0.0086	0.0121	-0.1501	0.0121	0.0346
400	0.0286	0.0120	0.0129	0.0362	0.0038	0.0051	-0.1381	0.0058	0.0249
800	0.0222	0.0044	0.0049	0.0244	0.0012	0.0018	-0.1146	0.0034	0.0165
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0338	0.0259	0.0270	0.0611	0.0093	0.0130	-0.1517	0.0127	0.0357
400	0.0223	0.0138	0.0143	0.0375	0.0040	0.0054	-0.1394	0.0065	0.0259
800	0.0185	0.0049	0.0053	0.0237	0.0015	0.0021	-0.1149	0.0031	0.0163

Next, to see how an increase in the variation of individual characteristics could impact the estimator performance, we consider a design where x_i is uniformly distributed over $\{0, 1, 2\}$. As in Table 1-3, we let the link formation probability to diminish as $n \to \infty$. For simplicity in implementation, we set link formation probability as follows: $nP_n\{g_{ij} = 1|x_i = x_j = 0\} = 10$, $nP_n\{g_{ij} = 1|x_i + x_j = 1\} = 8$, $nP_n\{g_{ij} = 1 \mid |x_i - x_j| = 2\} = 5$, $nP_n\{g_{ij} = 1|x_i = x_j = 1\} = 12$, $nP_n\{g_{ij} = 1|x_i + x_j = 3\} = 8$ and $nP_n\{g_{ij} = 1|x_i = x_j = 2\} = 10$ for all n.

Table 4 reports the results for such a design with m_i defined as the censored number of same-type neighbors. Compared with Table 3, the MSEs for β_m and τ, or the contextual and endogenous effects of m_i, are both higher while that for β_x, the marginal effect of individual characteristics, is slightly smaller. We interpret such a pattern as the result of an interaction of two immediate consequences of a larger support of X: on the one hand, a richer support for individual characteristics provides more sources of variation for recovering the parameter; on the other hand this increases the curse of dimensionality in that for a given sample size n there are fewer observations (individuals) that can be used to estimate the asymptotic moments in the first step (both of which condition on individual characteristics).

Table 5: Discretized m_i; $X = \{0, 1\}$; fixed $p = (0.6, 0.4, 0.6)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0286	0.0572	0.0580	0.2388	1.0166	1.0736	0.1810	0.0510	0.0838
400	0.0178	0.0094	0.0097	-0.1048	0.6867	0.6977	0.0980	0.0263	0.0359
800	0.0080	0.0045	0.0046	-0.0083	0.2713	0.2782	0.0260	0.0177	0.0184
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0475	0.0940	0.0963	-0.0622	2.2340	2.2378	0.2771	0.0826	0.1593
400	-0.0300	0.0140	0.0149	-0.0117	1.1734	1.1736	0.1503	0.0467	0.0692
800	-0.0155	0.0090	0.0092	0.0053	0.8007	0.8007	0.0201	0.0303	0.0307

Up to now we have only considered the designs where the link formation probabality varies with the sample size n. The last two tables of this section, Tables 5 and 6, report the simulation results in designs where the links are conditionally independently formed with fixed probability that is invariant with n. In both designs, the neighbor profiles are defined as the discretized average neighbor characteristic (as in Table 1). The two designs differ in the support of x_i.

The main takeaways from Tables 5 and 6 are as follows. These tables demonstrate evidence of convergence of MSEs for all three estimators under a fixed probability design. While the estimation error for $\hat{\beta}_x$ and $1 - \hat{\gamma}$ is comparable to their counterparts under convergent link formation probability (in Table 1), the MSEs for $\hat{\beta}_m$ are greater. We conjecture that this distinction happens because an increasing sample size n has different implication for the distribution of neighbor profile m_i under two paradigms of fixed or convergent link formation probabilities.

Table 6: Discretized m_i; $X = \{0, 1, 2\}$; fixed $p = (0.6, 0.4, 0.3, 0.6, 0.4, 0.6)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0124	0.0250	0.0251	-0.0646	0.8372	0.8414	0.3398	0.0312	0.1467
400	0.0108	0.0026	0.0027	-0.0754	0.4497	0.4554	0.1202	0.0077	0.0221
800	-0.0046	0.0012	0.0012	0.0312	0.3069	0.3079	-0.0633	0.0053	0.0093
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$1-\hat{\gamma}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	-0.0206	0.0265	0.0269	-0.1845	1.2083	1.2423	0.3431	0.0335	0.1512
400	-0.0113	0.0057	0.0058	-0.0968	0.8013	0.8106	0.1403	0.0128	0.0325
800	0.0045	0.0025	0.0025	0.0191	0.6374	0.6377	-0.1225	0.0092	0.0242

6 Concluding Remarks

Directions for future research include: prediction of counterfactual outcome when peer weights are assigned differently or higher-order neighbors are allowed to impact individual incentives; the use of higher moments of individual actions in identification and estimation; estimation when state variables are continuous; and richer models where network formation is endogenized along with individual actions.

The model and estimation method we propose in this article can be used to analyze individual incentives in a variety of environments. Examples include individual or household choices of consumption or investment levels in large social networks. Yet another interesting direction for future research would be the empirical analyses of individual preferences and interactions under such scenarios.

REFERENCES

Auerbach, E. "Identification and Estimation of a Partially Linear Regression Model using Network Data". Econometrica, Vol 91, Issue 1, page 347-365

Badev, A. "Discrete Games with Endogenous Networks: Theory and Policy," Econometrica, 2021.

Bajari, P., Hong, H., Krainer, J. and D. Nekipelov. "Estimating Static Models of Strategic Interactions." Journal of Business and Economic Statistics, 2010, Vol 28 (4), p.469-482

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. Jackson. "Gossip: Identifying Central Individuals in a Social Network," Working paper, NBER, 2014.

Blume, L., W. Brock, S. Durlauf, and R. Jayaraman. "Linear Social Interactions Models," Journal of Political Economy, 2015, 123 (2), 444-496.

Boucher, V. and I. Mourifié. My Friend Far, Far Away: A Random Field Approach to Exponential Random Graph Models. The Econometrics Jouranl, Vol 20, Issue 3, 2017

Bramoullé, Y., H. Djebbari, and B. Fortin. "Identification of Peer Effects through Social Networks," Journal of Econometrics, 2009, 150 (1), 44-55.

Canen, N., Schwartz, J. and K. Song. "Estimating Local Interactions Among Many Agents Who Observe Their Neighbors," Quantitative Economics, 2020, Vol 11, page 917-956.

Chandrasekhar, A. "Econometrics of Network Formation," Oxford Handbook on the Economics of Networks, 2016.

Chandrasekhar, A. and M. Jackson. "A Network Formation Model based on Subgraphs," Stanford working paper, 2021.

Christakis, N., J. Fowler, G. Imbens and K. Kalyanaraman. "An Empirical Model for Strategic Network Formation," The Econometric Analysis of Network Data. 2020. Elsevier.

De Giorgi, G., A. Frederiksen and L. Pistaferri. "Consumption Network Effects," Review of Economic Studies, forthcoming.

De Paula, A., S. Richards-Shubik and E. Tamer. "Identification of Preferences in Network

Formation Games," Econometrica, 86 (1), January 2018, pp.263-288

Galeotti, A., S. Goyal, M. Jackson, F. Vega-Redondo, and L. Yariv. "Network Games," Review of Economic Studies, 2010, 77 (1), 218-244.

Caldarelli, G., A. Capocci, P. De Los Rios, and M. A. Muñoz. "Scale-Free Networks from Varying Vertex Instrinsic Fitness." Physical Review Letters, 2002, Vol 89, No. 25

Giorgi, G., Frederiksen, A. and L. Pistaferri. "Consumption Network Effects," Review of Economic Studies, 2020, Vol 87, p.130-161

Hsieh, C., L. Lee, and V. Boucher. "Specification and estimation of network formation and network interaction models with the exponential probability distribution". Quantitative Econoimcs, 2020, Vol 11, Issue 4, page 1349-1390

Hotz, J. and R. Miller. "Conditional Choice Probabilities and the Estimation of Dynamic Models," Review of Economic Studies, 1993, vol.60, issue 3, 497-529.

Kojevnikov, D. and K. Song, "Econometric Inference on a Large Bayesian Game with Heterogeneous Beliefs," The Journal of Econometrics, forthcoming, 2023.

Le Cam, L. "An Approximation Theorem for the Poisson Binomial Distribution". Pacific Journal of Mathematics, Vol 10, Issue 4, page 1181-1197

Lee, L., Liu, X. and X. Lin. "Specification and Estimation of Social Interaction Models with Network Structures," The Econometrics Journal, 2010, Vol. 13 (2), p.145-176.

Leung, M., "Two-Step Estimation of Network-Formation Models with Incomplete Information," Journal of Econometrics, 2015, 188 (1), 182-195.

Leung, M., "Dependence-robust Inference using Resampled Statistics," Journal of Applied Econometrics, 2021, Vol 37 (2), 270-285.

Leung, M. and H. R. Moon. "Normal Approximation in Large Network Models," Working paper, ArXiv, 2023.

Lewbel, A. "The Identification Zoo: Meanings of Identification in Econometrics". Journal of Economic Literature, 2019, Vol 57, Issue 4, page 835-903

Lin, Z. and H. Xu, "Estimation of Social-influence-dependent Peer Pressure in a Large

Network Game," Econometrics Journal, 2017.

Manski, C., "Identification of Endogenous Social Effects: The Reflection Problem," Review of Economic Studies, 1993, 60 (3), 531-542.

Mele, A., "A Structural Model of Segregation in Social Networks," Econometrica, forthcoming, 2017. Vol 85, Issue 3.

Menzel, K., "Inference for Games with Many Players," Review of Economic Studies, 2015, 88 (1), 306-337.

Miyauchi, Y. "Structural Estimation of a Pairwise Stable Network with Nonnegative Externality," Journal of Econometrics, 2016, 195 (2), 224-235.

Newey, W. and D. McFadden, "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. New York: Elsevier, 1994.

Powell, J. L. "Estimation of Semiparametric Models," *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. New York: Elsevier, 1994.

Sheng, S., "A Structural Econometric Analysis of Network Formation Games," Econometrica, 2020. Vol 88, Issue 5.

Todd, P. and K. Wolpin, "Accounting for Mathematics Performance of High School Students in Mexico: Estimating a Coordination Game in the Classroom," Journal of Political Economy, 2018, Vol. 126, No. 6, p.2608-2650

Xu, H., "Social Interactions on Large Networks: A Game Theoretic Approach," International Economic Review, 2018.

Xu, X. and L. Lee, "Estimation of a Binary Choice Game Model with Network Links," Working paper, Ohio State University, 2015.

Yang, C. and L. Lee, "Social Interactions under Incomplete Information with Heterogeneous Expectations," Journal of Econometrics, 2017, Volume 198, Issue 1

Appendix

A. Proofs in Section 3

Proof of Theorem 1. Let $\mathcal{S}$ denote the set of bounded functions on T^n taking values in A^n, where T denotes the support of τ_i as in the text. (The boundedness condition can be dispensed; the proof applies when $\mathcal{S}$ is a complete metric space.) For $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$, let $\mathbf{s} \leq \mathbf{s}'$ denote $\mathbf{s}(\tau) \leq \mathbf{s}'(\tau)$ for all $\tau \in T^n$. For any $\mathbf{s} \in \mathcal{S}$, let $\|\mathbf{s}\| = \max_i \sup\{|s_i(\tau_i)| : \tau_i \in T\}$, i.e., $\|.\|$ is the supremum norm. Note $\mathcal{S}$ with the supremum norm is a complete metric space.

Define a mapping $\mathbf{R} : \mathcal{S} \to \mathcal{S}$ as

$$\mathbf{R}(\mathbf{s})_i(\tau_i) \equiv R_i(\tau_i; \mathbf{s}_{-i}) \text{ for all } i \in N.$$

First note that $\mathbf{R}(\mathbf{s}) \in \mathcal{S}$ for any $\mathbf{s} \in \mathcal{S}$, and so $\mathbf{R}$ maps $\mathcal{S}$ to itself. To establish existence of a unique p.s.BNE, we show that $\mathbf{R}$ satisfies the contraction property. That is, there exists $c \in (0,1)$ such that $\|\mathbf{R}(\mathbf{s}) - \mathbf{R}(\mathbf{s}')\| \leq c\|\mathbf{s} - \mathbf{s}'\|$ for any $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$. An application of the Banach Fixed Point Theorem then proves the existence of a unique p.s.BNE.

To show that $\mathbf{R}$ satisfies the contraction property, fix any $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$. Note that for any $i \in N$ and $\tau_i \in T$, we have:

$$\begin{aligned} |R_i(\tau_i; \mathbf{s}_{-i}) - R_i(\tau_i; \mathbf{s}'_{-i})| &= \frac{\phi \sum_{j \in N_i} \tilde{w}_{ij}(\tau_i) \left| \mathbf{E}\left[s_j(\tau_j) - s'_j(\tau_j) \middle| \tau_i \right] \right|}{1 + \phi \sum_{j \in N_i} \tilde{w}_{ij}(\tau_i)} \\ &\leq (1 - \tilde{\gamma})\|\mathbf{s} - \mathbf{s}'\|, \end{aligned} \tag{12}$$

for a constant $\tilde{\gamma} \in (0,1)$, where the inequality is due to the sup norm, $\phi > 0$ and boundedness of postive peer effect weights $\tilde{w}_{ij}(\tau_i)$. By definition, $\|\mathbf{R}(\mathbf{s}) - \mathbf{R}(\mathbf{s}')\| = \sup\{|R_i(\tau_i; \mathbf{s}_{-i}) - R_i(\tau_i; \mathbf{s}'_{-i})| : \tau_i \in T\}$ for some $i \in N$. Hence the desired result follows from (12). $\square$

Proof of Proposition 1. Under Assumptions 1, 2 and 3, a unique symmetric pure-strategy Bayesian Nash equilibrium exists in each data-generating process indexed by n, and

$$a_i = s(t_i) = \gamma h(t_i) + (1 - \gamma) \sum_{x' \in X} w(x', x_i, m_i) \mathbf{E}_n[s(t_j) | x_j = x', x_i, m_i, g_{ij} = 1].$$

Define $\lambda_n(x, m) \equiv \frac{1}{n} \sum_i \mathbf{E}_n(a_i | x_i = x, m_i = m) = \mathbf{E}_n(a_i | x_i = x, m_i = m)$, which does not vary with the specific identity of an individual i because of the symmetry in Assumption 2 and exchangeability in Assumption 3. Likewise, let $\bar{h}_n(x,m) \equiv \frac{1}{n}\sum_i \mathbf{E}_n\left[h(t_i)|x_i = x, m_i = m\right] = \mathbf{E}_n\left[h(t_i)|x_i = x, m_i = m\right]$. Thus by construction,

$$\lambda_n(x,m) = \gamma\bar{h}_n(x,m) + (1-\gamma)\sum_{x'\in X} w(x',x,m)\mathbf{E}_n[s(t_j)|x_j = x', m_i = m, x_i = x, g_{ij} = 1]. \tag{13}$$

For any n, the law of total expectation and Assumption 4 imply

$$\begin{aligned}
& \mathbf{E}_n[s(t_j)|x_j = x', x_i = x, m_i = m, g_{ij} = 1] \\
= & \sum_{m'\in M} \mathbf{E}_n(a_j|m_j = m', x_j = x')\mathbf{E}_n(1\{m_j = m'\}|x_j = x', x_i = x, m_i = m, g_{ij} = 1) \\
= & \sum_{m'\in M} \lambda_n(x',m')q_n(m'|x',m,x),
\end{aligned} \tag{14}$$

where $q_n(m'|x',x,m) \equiv \frac{1}{n(n-1)}\sum_{j\neq i}\mathbf{E}_n(1\{m_j = m'\}|x_j = x', x_i = x, m_i = m, g_{ij} = 1)$. The second equality holds because $\mathbf{E}_n(1\{m_j = m'\}|x_j = x', x_i = x, m_i = m, g_{ij} = 1)$ does not depend on specific identities of individuals i and j under Assumptions 2 and 3. Combining (13) and (14), we write λ_n as the solution to a fixed-point equation that depends on $(\bar{h}_n, q_n)$. That is,

$$\lambda_n = \Gamma(\lambda_n; \bar{h}_n, q_n), \tag{15}$$

where $\Gamma(.;\bar{h}_n, q_n)$ is a self-map over the set of bounded and continuous functions with domain $X\times M$, and

$$\begin{aligned}
\Gamma(\lambda_n; \bar{h}_n, q_n)(x,m) \equiv & \ \gamma\bar{h}_n(x,m) \\
& + \ (1-\gamma)\textstyle\sum_{x'\in X} w(x',x,m)\sum_{m'\in M}\lambda_n(x',m')q_n(m'|x',x,m).
\end{aligned}$$

The solution is unique because, for any $\bar{h}_n$ and q_n, the map $\Gamma(.;\bar{h}_n, q_n)$ has a contraction property under the sup norm.[12]

Next, let h and $\tilde{h}$ denote generic functions of (x_i, m_i) determining the contextual effects for i; let q and $\tilde{q}$ denote generic density (probability mass) functions of m' given x', x, m. We now show that for any $\bar{c} > 0$ there exists $c_1, c_2 > 0$ so that

$$\|\tilde{h} - h\| \leq c_1 \text{ and } \|\tilde{q} - q\| \leq c_2 \text{ imply } \|\tilde{\lambda} - \lambda\| \leq \bar{c}, \tag{16}$$

[12]The proof of the contraction property of $\Gamma(.;\bar{h}_n, q_n)$ is similar to that of Theorem 1, and omitted for brevity.

where $\|.\|$ denotes the sup norm over the respective domains, and λ and $\tilde{\lambda}$ are the unique solutions in $\lambda = \Gamma(\lambda; h, q)$ and $\tilde{\lambda} = \Gamma(\tilde{\lambda}; \tilde{h}, \tilde{q})$ respectively. To verify (16), recursively substitute λ and $\tilde{\lambda}$ into $\Gamma(\lambda; h, q)$ and $\Gamma(\tilde{\lambda}; \tilde{h}, \tilde{q})$ and use the fact that "$\gamma \in (0,1)$, $w(., x, m) \geq 0$ and $\sum_{x' \in X} w(x', x, m) = 1$ for all $x \in X, m \in M$".

It then follows from (16) that the solution to the fixed point problem $\lambda = \Gamma(\lambda; h, q)$ is continuous in (h, q). Under Assumption 5, $\bar{h}_n \to h^*$ and $q_n \to q^*$ given the sup norm. Consequently, the sequence of solutions $\lambda_n = \Gamma(\lambda_n; \bar{h}_n, q_n)$ converges to the unique λ^* that solves $\lambda^* = \Gamma(\lambda^*; h^*, q^*)$. $\square$

Proof of Proposition 2. Fix some positive integer $n \in \mathbb{N}_{++}$, $x \in X$ and $m \in M$. The Chebychev's Inequality implies that for any constant $c > 0$,

$$\mathbf{P}_n \left\{ \left| \frac{1}{n} \sum\nolimits_i \iota_i - \mathbf{E}_n \left(\frac{1}{n} \sum\nolimits_i \iota_i \right) \right| \geq c \right\} \leq c^{-2} \mathbf{V}_n \left(\frac{1}{n} \sum\nolimits_i \iota_i \right), \tag{17}$$

where $\mathbf{P}_n$ is the probability measure associated with F_n, and $\mathbf{V}_n$ denotes the variance under F_n. In what follows, let $\sigma_{n,i}^2$ be a shorthand for $\mathbf{V}_n(\iota_i)$; and let $\mathbf{C}_{n,i,j}$ be a shorthand for $\mathbf{C}_n(\iota_i, \iota_j)$, which is the covariance between ι_i and ι_j under F_n. By the exchangeability and anonymity of F_n in Assumption 3, $\sigma_{n,i}$ does not depend on i and $\mathbf{C}_{n,i,j}$ does not depend on i and j. Therefore the right-hand side of (17) can be written as

$$c^{-2} n^{-2} \left[\sum\nolimits_i \sigma_{n,i}^2 + \sum\nolimits_{j \neq i} \mathbf{C}_{n,i,j} \right] = \frac{n \sigma_{n,i}^2}{n^2 c^2} + \frac{n(n-1) \mathbf{C}_{n,i,j}}{n^2 c^2}. \tag{18}$$

By the first asymptotic uncorrelation condition in Assumption 6, the two terms on the right-hand of (18) converge to 0 as $n \to \infty$. Thus $\frac{1}{n} \sum_i \iota_i - \frac{1}{n} \sum_i \mathbf{E}_n(\iota_i) \xrightarrow{p} 0$ as $n \longrightarrow \infty$. Under Assumption 5, this implies $\frac{1}{n} \sum_i \iota_i(x, m) \xrightarrow{p} p^*(x, m)$ as $n \longrightarrow \infty$.

Next, recall that a_i is a function of $(x_i, m_i, \varepsilon_i)$ in p.s.BNE. By the law of total covariance, $\mathbf{C}_n(a_i \iota_i(x, m), a_j \iota_j(x, m)) \to 0$ under conditional independence in Assumption 4 and the first asymptotic uncorrelation condition in Assumption 6. It follows from a similar argument using Chebychev's Inequality that $\frac{1}{n} \sum_i a_i \iota_i - \frac{1}{n} \sum_i \mathbf{E}_n(a_i \iota_i) \xrightarrow{p} 0$. Under Assumption 5 and Proposition 1, the limit of $\frac{1}{n} \sum_i \mathbf{E}_n(a_i \iota_i)$ as $n \to \infty$ exists. By exchangeability of F_n in Assumption 3 both $\mathbf{E}_n(\iota_i)$ and $\mathbf{E}_n(a_i \iota_i)$ do not vary across the identities of individuals i. It then follows from the Slutsky's Theorem that

$$\frac{\frac{1}{n} \sum_i a_i \iota_i(x, m)}{\frac{1}{n} \sum_i \iota_i(x, m)} \xrightarrow{p} \frac{\lim_{\tilde{n} \to \infty} \mathbf{E}_{\tilde{n}}[a_i \iota_i(x, m)]}{\lim_{\tilde{n} \to \infty} \mathbf{E}_{\tilde{n}}[\iota_i(x, m)]} = \lim_{\tilde{n} \to \infty} \frac{\mathbf{E}_{\tilde{n}}[a_i \iota_i(x, m)]}{\mathbf{E}_{\tilde{n}}[\iota_i(x, m)]} = \lambda^*(x, m) \tag{19}$$

for all x, m on the support of p^*.

We now prove the second claim in the proposition. Fix $x, x' \in X$ and $m, m' \in M$. In what follows, let ι'_j, ι_i be shorthand for $\iota_j(x', m'), \iota_i(x, m)$ respectively. Define $\xi_{ij} \equiv (\iota'_j \iota_i g_{ij} + \iota'_i \iota_j g_{ji})/2$ so that $\xi_{ij} = \xi_{ji}$ and $\frac{1}{n(n-1)} \sum_{j \neq i} \iota'_j \iota_i g_{ij} = \frac{2}{n(n-1)} \sum_i \sum_{j>i} \xi_{ij}$ by construction. By the Chebychev Inequality, for any constant $c > 0$,

$$\mathbf{P}_n \left\{ \left| \tfrac{2}{n(n-1)} \sum\nolimits_{j \neq i} \xi_{ij} - \mathbf{E}_n \left(\tfrac{2}{n(n-1)} \sum\nolimits_{j \neq i} \xi_{ij} \right) \right| \geq c \right\} \leq c^{-2} \mathbf{V}_n \left(\tfrac{2}{n(n-1)} \sum\nolimits_{j \neq i} \xi_{ij} \right),$$

where the right-hand side is

$$\tfrac{4}{c^2 n^2 (n-1)^2} \sum\nolimits_{j>i} \sum\nolimits_{t>l} \mathbf{C}_n \left(\xi_{ij}, \xi_{lt} \right).$$

By construction, this quadruple sum consists of $\binom{n}{2} \times \binom{n}{2} = \frac{1}{4}(n^4 - 2n^3 + n^2)$ terms. These include $\binom{n}{2} = \frac{1}{2}(n^2 - n)$ variance terms $\mathbf{V}_n(\xi_{ij})$, $\binom{n}{2} \times \binom{n-2}{2} = \frac{1}{4}(n^4 - 6n^3 + 11n^2 - 6n)$ covariance terms $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ in which the unordered pairs $\{i, j\}$ and $\{l, t\}$ do not overlap, and $\binom{n}{2} \times \left(\binom{n}{2} - \binom{n-2}{2} - 1 \right) = n^3 - 3n^2 + 2n$ covariance terms $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ in which the two pairs $\{i, j\}$ and $\{l, t\}$ share *exactly one* individual in common. Note that $\mathbf{V}_n(\xi_{ij})$ and $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ are bounded for all $\{i, j\}$ and $\{l, t\}$. Furthermore, the covariance term $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ with $\{i, j\} \cap \{l, t\} = \emptyset$ does not vary with the identities $\{i, j, l, t\}$ due to the anonymity of common prior in Assumption 3. Under the asymptotic uncorrelation condition in Assumption 6, $\mathbf{C}_n(\xi_{ij}, \xi_{lt}) \to 0$ as $n \to \infty$ if $\{i, j\} \cap \{l, t\} = \emptyset$. Therefore $\frac{4}{c^2 n^2 (n-1)^2} \sum_{j>i} \sum_{l>t} \mathbf{C}_n(\xi_{ij}, \xi_{lt}) \to 0$ as $n \to \infty$. Hence $\frac{1}{n(n-1)} \sum_{j \neq i} [\iota'_j \iota_i g_{ij} - \mathbf{E}_n(\iota'_j \iota_i g_{ij})] \xrightarrow{p} 0$. By a similar argument, $\frac{1}{n(n-1)} \sum_{j \neq i} [1\{x_j = x'\} \iota_i g_{ij} - \mathbf{E}_n(1\{x_j = x'\} \iota_i g_{ij})] \xrightarrow{p} 0$. Under our condition in Assumption 5, $\lim_{\tilde{n} \to \infty} \mathbf{E}_{\tilde{n}}(\iota'_j \iota_i g_{ij})$ exists and $\lim_{\tilde{n} \to \infty} \mathbf{E}_{\tilde{n}}(1\{x_j = x'\} \iota_i g_{ij})$ is non-zero. The second convergence result in the proposition follows from an argument analogous to (19) under the exchangeability in Assumption 3 and the existence of the limits q^* in Assumption 5. □

Proof of Proposition 3. Suppose (θ, γ, ω) is observationally equivalent to some different vector of parameters $(\tilde{\theta}, \tilde{\gamma}, \tilde{\omega})$ based on asymptotic moments in (11). This means (11) holds almost surely p^* when (γ, θ, ω) is replaced by $(\tilde{\gamma}, \tilde{\theta}, \tilde{\omega})$. For each $d \in \Phi$, let $\omega(d) \equiv (\omega(c, d))_{c \in \Psi}$, which is a column-vector of weights assigned over Ψ conditional on $\varphi(x, m) = d$. Likewise, define $\tilde{\omega}$ using $\tilde{\omega}(.,.)$. That is, for every $d \in \Phi$,

$$\gamma \eta(x, m; \theta) + (1 - \gamma) \mu^*(x, m) \omega(d) = \tilde{\gamma} \eta(x, m; \tilde{\theta}) + (1 - \tilde{\gamma}) \mu^*(x, m) \tilde{\omega}(d) \tag{20}$$

whenever $\varphi(x,m) = d$.

Consider the following cases. Case (i): $\theta = \tilde{\theta}$ and $(\gamma,\omega) \neq (\tilde{\gamma},\tilde{\omega})$. Then (20) implies that for each $d \in \Phi$, $[\eta(x,m;\theta), \mu^*(x,m)]\alpha(d) = 0$ whenever $\varphi(x,m) = d$, where $\alpha(d) \equiv [\gamma - \tilde{\gamma}, (1-\gamma)\omega(d)' - (1-\tilde{\gamma})\tilde{\omega}(d)']'$. Because $(\gamma,\omega) \neq (\tilde{\gamma},\tilde{\omega})$, the vector $\alpha(d)$ must be nonzero at least for some $d \in \Phi$. This implies that *at least for some* $d \in \Phi$, $[\eta(x,m;\theta), \mu^*(x,m)]$ does not have full rank conditional on $\varphi(x,m) = d$. Case (ii): $\theta \neq \tilde{\theta}$ and $(\gamma,\omega) = (\tilde{\gamma},\tilde{\omega})$. Then (20) implies $\eta(x,m;\theta) = \eta(x,m;\tilde{\theta})$ almost everywhere p^*. Case (iii): $\theta \neq \tilde{\theta}$ and $(\gamma,\omega) \neq (\tilde{\gamma},\tilde{\omega})$. Then (20) implies that for every $d \in \Phi$,

$$[\eta(x,m;\theta), \eta(x,m;\tilde{\theta}), \mu^*(x,m)]\mathbf{b}(d) = 0$$

whenever $\varphi(x,m) = d$, where $\mathbf{b}(d) \equiv [\gamma, -\tilde{\gamma}, (1-\gamma)\omega(d)' - (1-\tilde{\gamma})\tilde{\omega}(d)']'$. By construction, $\mathbf{b}(d)$ is non-zero for all d. Thus (20) implies that *for each* $d \in \Phi$, $[\eta(x,m;\theta),\ \eta(x,m;\tilde{\theta}),\ \mu^*(x,m)]$ does not have full rank conditional on $\varphi(x,m) = d$.

Each of these cases of observational equivalence implies the following condition: "There exists $\tilde{\theta}$ such that *at least for some* $d \in \Phi$, $[\eta(x,m;\theta),\ \eta(x,m;\tilde{\theta}),\ \mu^*(x,m)]$ does not have full rank conditional on $\varphi(x,m) = d$ under p^*". It then follows that under Assumption 9, (θ,γ,ω) is not observationally equivalent to any $(\tilde{\theta},\tilde{\gamma},\tilde{\omega}) \neq (\theta,\gamma,\omega)$. □

Proof of Proposition 4. Suppose (θ,γ,ω) is observationally equivalent to some other $(\tilde{\theta},\tilde{\gamma},\tilde{\omega})$. This implies that for each $d \in \Phi$, $[\zeta_d(x,m), \mu^*(x,m)]\alpha(d) = 0$ when $\varphi(x,m) = d$, where $\alpha(d) \equiv [\gamma\theta_d - \tilde{\gamma}\tilde{\theta}_d, (1-\gamma)\omega(d)' - (1-\tilde{\gamma})\tilde{\omega}(d)']'$. Consider two cases. Case (i): $\omega = \tilde{\omega}$. In this case, either $\gamma \neq \tilde{\gamma}$ or $\theta_d \neq \tilde{\theta}_d$ at least for some $d \in \Phi$. Otherwise (θ,γ,ω) would be identical to $(\tilde{\theta},\tilde{\gamma},\tilde{\omega})$. Hence at least for some d, the two terms $\gamma\theta_d - \tilde{\gamma}\tilde{\theta}_d$ and $(1-\gamma)\omega(d)' - (1-\tilde{\gamma})\tilde{\omega}(d)'$ can not be zero simultaneously. Thus $\alpha(d)$ is a non-zero vector for all d. This implies that *at least for some* $d \in \Phi$, $[\zeta_d(x,m), \mu^*(x,m)]$ does not have full rank conditional on $\varphi(x,m) = d$. Case (ii): $\omega \neq \tilde{\omega}$. In this case, $\omega(d) \neq \tilde{\omega}(d)$ at least for some d. It then follows that $(1-\gamma)\omega(d) - (1-\tilde{\gamma})\tilde{\omega}(d)$ is a non-zero vector at least for some d, regardless of whether $\gamma \neq \tilde{\gamma}$. This implies that *at least for some* $d \in \Phi$, $[\zeta_d(x,m), \mu^*(x,m)]$ does not have full rank conditional on $\varphi(x,m) = d$. Therefore, if $[\zeta_d(x,m), \mu^*(x,m)]$ has full rank conditional on $\varphi(x,m) = d$ for all $d \in \Phi$, then (θ,γ,ω) is not observationally equivalent to any other $(\tilde{\theta},\tilde{\gamma},\tilde{\omega})$. □

Proof of Proposition 5. Let $\sum_{x,m}$, $\sum_k$, $\sum_\ell$ be shorthand for $\sum_{x\in X, m\in M}$, $\sum_{k=1,2}$, $\sum_{\ell=1,2}$ respectively. For each $\delta \in D$, let

$$G_0(\delta) \equiv \sum_{x,m} p^*(x,m)\left[\lambda^*(x,m) - \gamma\eta(x,m;\theta) - \sum_k 1\{z = z_{(k)}\}\left(\sum_\ell \chi^*_\ell(x,m)\beta_{k\ell}\right)\right]^2,$$

where

$$\chi^*_\ell(x,m) \equiv \tfrac{1}{\#V}\sum_{v',m'} \lambda^*(v', z_{(\ell)}, m')q^*(m'|v', z' = z_{(\ell)}, x, m).$$

Define $\tilde{G}_n(\delta) \equiv n^{-1}\sum_i \Upsilon_i(\delta) \equiv \overline{\Upsilon}_n(\delta)$, with

$$\Upsilon_i(\delta) \equiv \Upsilon(x_i, m_i; \delta) \equiv [\lambda^*_i - \gamma\eta_i(\theta) - \sum_k 1\{z_i = z_{(k)}\}\sum_l \chi^*_{l,i}\beta_{kl}]^2,$$

where $\lambda^*_i \equiv \lambda^*(x_i, m_i)$ and $\chi^*_{l,i} \equiv \chi^*_l(x_i, m_i)$. For every fixed δ, apply a first-order Taylor expansion of $\hat{G}_n(\delta)$ around λ^*_i and $\chi^*_{l,i}$. By bounded asymptotic moments in T3-(i) and Proposition 2, $\hat{G}_n(\delta) - \tilde{G}_n(\delta) \xrightarrow{p} 0$ for all δ. Next, by the Chebychev's Inequality,

$$\mathbf{P}_n\left\{\left|\overline{\Upsilon}_n(\delta) - \mathbf{E}_n\left[\overline{\Upsilon}_n(\delta)\right]\right| \geq c\right\} \leq c^{-2}\mathbf{V}_n\left[\overline{\Upsilon}(\delta)\right] \text{ for any constant } c > 0.$$

For any fixed δ, let $\tilde{\sigma}^2_{n,i}, \tilde{C}_{n,i,j}$ be shorthand for $\mathbf{V}_n[\Upsilon_i(\delta)]$, $\mathbf{C}_n[\Upsilon_i(\delta), \Upsilon_j(\delta)]$ respectively. By the exchangeability and anonymity of F_n in Assumption 3, $\tilde{\sigma}_{n,i}, \tilde{C}_{n,i,j}$ are invariant in the subscripts i, j. Hence, the right-hand of the Chebychev's inequality above is:

$$c^{-2}n^{-2}\left[\sum_i \tilde{\sigma}^2_{n,i} + \sum_{j\neq i}\tilde{C}_{n,i,j}\right] = \frac{n\tilde{\sigma}^2_{n,i}}{c^2n^2} + \frac{n(n-1)\tilde{C}_{n,i,j}}{c^2n^2}.$$

With $\tilde{\sigma}^2_{n,i} < \infty$, the first term on the right-hand side converges to zero as $n \to \infty$. Recall that, under Assumption 6, $1\{(x_i, m_i) = (x,m)\}$ and $1\{(x_j, m_j) = (x,m)\}$ are asymptotically uncorrelated for all values (x,m). Hence the arguments in $\Upsilon(\cdot,\cdot;\delta)$, i.e., (x_i, m_i), are asymptotically independent across individuals i, j. Consequently, the second term on the right-hand side also converges to zero as $n \to \infty$. It then follows that $\overline{\Upsilon}_n(\delta) - \mathbf{E}_n\left[\overline{\Upsilon}_n(\delta)\right] \xrightarrow{p} 0$ for all δ. Next, note:

$$\begin{aligned}
\mathbf{E}_n\left[\overline{\Upsilon}_n(\delta)\right] &= \mathbf{E}_n\left[n^{-1}\sum_i\sum_{(x,m)\in X\times M} 1\{(x_i,m_i) = (x,m)\} \times \Upsilon(x,m;\delta)\right] \\
&= \sum_{(x,m)\in X\times M}\left\{n^{-1}\sum_i \mathbf{E}_n\left[1\{(x_i,m_i) = (x,m)\}\right]\right\} \times \Upsilon(x,m;\delta) \\
&\to \sum_{(x,m)\in X\times M} p^*(x,m) \times \Upsilon(x,m;\delta) \equiv G_0(\delta),
\end{aligned}$$

where the summation $\sum_{(x,m)\in X\times M}$ is over all values in the join support of (x_i, m_i), and the convergence follows from the definition of p^* in Assumption 5.

For notational convenience, let $1_{k,i} \equiv 1\{z_i = z_{(k)}\}$,

$$\hat{\mathbf{H}}_i \equiv [1_{1,i}\hat{\chi}_{1,i}, 1_{1,i}\hat{\chi}_{2,i}, 1_{2,i}\hat{\chi}_{1,i}, 1_{2,i}\hat{\chi}_{2,i}]$$

and

$$\bar{q}_i(\theta) \equiv \hat{\lambda}_i - \gamma\eta_i(\theta) - \sum\nolimits_k \mathbf{1}_{k,i} \sum\nolimits_\ell \hat{\chi}_{\ell,i}\beta_{k\ell}.$$

For any $\delta \neq \delta'$, the mean value approximation implies $\hat{G}_n(\delta') - \hat{G}_n(\delta) = \nabla_\delta \hat{G}_n(\tilde{\delta})(\delta' - \delta)$, where $\tilde{\delta}$ is an intermediate value between δ, δ' and the gradient $\nabla_\delta \hat{G}_n(\delta)$ is

$$\frac{2}{n}\sum\nolimits_i \bar{q}_i(\delta)[\eta_i(\theta), \hat{\mathbf{H}}_i, \gamma\nabla_\theta\eta_i(\theta)].$$

Let $\|.\|$ denote the Euclidean norm. By the Cauchy-Schwarz inequality,

$$|\nabla_\delta \hat{G}_n(\tilde{\delta})(\delta' - \delta)| \leq \left\|\nabla_\delta \hat{G}_n(\tilde{\delta})\right\| \times \|\delta' - \delta\|,$$

where $\left\|\nabla_\delta \hat{G}_n(\tilde{\delta})\right\|$ is $O_p(1)$ under T3. Then by Lemma 2.9 in Newey and McFadden (1994),

$$\sup_{\delta\in D} |\hat{G}_n(\delta) - G_0(\delta)| \xrightarrow{p} 0.$$

In addition, G_0 is continuous in δ over D under T3. Under the condition in T2, δ_0 is a unique maximizer of $G_0(.)$ over D. By Theorem 2.1 in Newey and McFadden (1994), $\hat{\delta}_n \xrightarrow{p} \delta_0$. $\square$

B. An Example of Random Network

This section presents an example of networks with *undirected* links (i.e., $g_{ij} = g_{ji}$ for any $i, j \in N$) that satisfy the conditions in Assumption 5 (existence of limits as $n \to \infty$) and Assumption 6 (asymptotic uncorrelation as $n \to \infty$).

As in Section 2, let N denote the set of individuals and let $n \equiv \#N \in \mathbb{N}_{++} \equiv \{1, 2, 3, .., \infty\}$. Each individual is characterized by a binary characteristic $x_i \in X \equiv \{1, 2\}$. (Generalization to the case $\#X \geq 3$ is straightforward.) Individual characteristics $x_i, i = 1, ..., n$ are independently drawn from a fixed multinomial distribution with $p_{(k)} \equiv \Pr\{x_i = k\}$ with $k = 1, 2$. Let $n_{(k)} \equiv \#\{i \in N : x_i = k\}$ for $k = 1, 2$. By an application of the Weak Law of Large Numbers, $n_{(k)} \to \infty$ and $n_{(k)}/n \to p_{(k)} \in (0, 1)$ as $n \to \infty$ for $k = 1, 2$. As in the text, let $\mathbf{E}_n(.)$ denote the expectation under F_n, or the distribution of $(\tau_i)_{i\in N_n}$ in the data-generating process (DGP) indexed by network size n, and let P_n denote the probability measure in the DGP. Consider a random Poisson network that satisfies the following conditions on link formation.

(R1). *For each n and $k, \ell = 1, 2$, $P_n\{g_{ij} = 1 | x_i = k, x_j = l\} = q_{(k\ell),n}$, where $q_{(k\ell),n} n_{(\ell)} \to \rho_{(k\ell)} < \infty$ as $n \to \infty$.*

(R2). *For each n, the distribution of links conditional on individual characteristics is*

$$\prod\nolimits_{j>i} \mathbf{E}_n(g_{ij}|x_i, x_j)^{\tilde{g}_{ij}} [1 - \mathbf{E}_n(g_{ij}|x_i, x_j)]^{1-\tilde{g}_{ij}},$$

where $\tilde{g}_{ij}$ denotes realized values of g_{ij}.

Under R2, the links are independent once conditional on the characteristics of individuals. Recall from Section 2 that an individual i's neighborhood profile is summarized by a vector of integers $\mathbf{n}_i \equiv (n_{i,1}, n_{i,2})$ with $n_{i,1} + n_{i,2} = \#N_i$, where $n_{i,k} \equiv \#\{j : g_{ij} = 1, x_j = k\}$ and $N_i \equiv \{j \in N : g_{ij} = 1\}$. Let $m_i \equiv (m_{i1}, m_{i2}) \equiv (\min\{n_{i,1}, \bar{n}_1\}, \min\{n_{i,2}, \bar{n}_2\})$. That is, for any fixed n, the k-th component in $m_i \in \mathbb{N}_+^2$ follows a binomial distribution censored at an exogenously fixed maximum number of type-k friends possible, denoted by $\bar{n}_k$ for $k = 1, 2$. By construction $M \equiv \{0, 1, ..., \bar{n}_1\} \times \{0, 1, ..., \bar{n}_2\}$ is finite and invariant as $n \to \infty$. For any n and any $\bar{m} \equiv (\bar{m}_1, \bar{m}_2) \in M$,

$$\begin{aligned} \mathbf{E}_n \left[1\{x_i = k, m_i = \bar{m}\}\right] &= p_{(k)} P_n\{m_{i1} = \bar{m}_1, m_{i2} = \bar{m}_2 | x_i = k\} \\ &= p_{(k)} P_n\{m_{i1} = \bar{m}_1 | x_i = k\} P_n\{m_{i2} = \bar{m}_2 | x_i = k\}, \end{aligned} \tag{21}$$

where the second equality follows from Condition R2. By definition, for any $\bar{m}_k < \bar{n}_k$,

$$\begin{aligned} P_n\{m_{ik} &= \bar{m}_k | x_i = k\} = P_n\{n_{ik} = \bar{m}_k | x_i = k\} \\ &= \binom{n_{(k)} - 1}{\bar{m}_k} \left[q_{(kk),n}\right]^{\bar{m}_k} \left[1 - q_{(kk),n}\right]^{n_{(k)} - \bar{m}_k - 1}. \end{aligned}$$

Under Condition R1, the Poisson Limit Theorem applies and the expression on the right-hand side converges to

$$\left[\rho_{(kk)}\right]^{\bar{m}_k} \exp\{-\rho_{(kk)}\} / (\bar{m}_k!),$$

which is the probability mass function (p.m.f.) of a random variable distributed as Poisson with mean $\rho_{(kk)}$. Furthermore,

$$P_n\{m_{ik} = \bar{n}_k | x_i = k\} = P_n\{n_{ik} \geq \bar{n}_k | x_i = k\}.$$

Under Condition R1, $q_{(k\ell),n} = O(n^{-1})$. This implies $nq^2_{(k\ell),n} \to 0$. Thus by the Le Cam's Theorem in Le Cam (1960) and an application of triangular inequality,

$$\left| P_n\{n_{ik} \geq \bar{n}_k | x_i = k\} - \sum\nolimits_{\bar{m}_k=\bar{n}_k}^{\infty} \left[\rho_{(kk)}\right]^{\bar{m}_k} \exp\{-\rho_{(kk)}\}/(\bar{m}_k!) \right| \to 0$$

as $n \to \infty$. That is, as the network size increases to infinity, $P_n\{n_{ik} \geq \bar{n}_k | x_i = k\}$ converges to the survival function (evaluated at $\bar{n}_k$) of a Poisson variable with mean $\rho_{(kk)}$. By a symmetric argument, we can show similar results with $l \neq k$: $P_n\{m_{il} = \bar{m}_\ell | x_i = k\}$ converges to a Poisson p.m.f. with mean $\rho_{(k\ell)}$ for $\bar{m}_\ell < \bar{n}_\ell$; and converges to the survival function at $\bar{n}_\ell$ of a Poisson variable with mean $\rho_{(k\ell)}$ for $\bar{m}_\ell = \bar{n}_\ell$. (To show this, replace $n_{(k)} - 1$, $q_{(kk),n}$, $\rho_{(kk)}$ in the argument above with $n_{(\ell)}$, $q_{(k\ell),n}$, $\rho_{(k\ell)}$ respectively, and apply the Le Cam's Theorem.) Thus the right-hand side of (21), and consequently $\mathbf{E}_n\left[1\{x_i = k, m_i = \bar{m}\}\right]$, converges to some non-zero limits $p^*(x, m)$ for all $\bar{m} \in M$ and $k = 1, 2$.

Consider an uncensored vectors $\tilde{m}$. Then

$$\begin{aligned} & \mathbf{E}_n(1\{m_j = \tilde{m}\} | x_j = 2, x_i = 1, m_i = \bar{m}, g_{ij} = 1) \\ = & \binom{n_{(1)} - 1}{\tilde{m}_1 - 1} \left[q_{(21),n}\right]^{\tilde{m}_1 - 1} \left[1 - q_{(21),n}\right]^{n_{(1)} - \tilde{m}_1} \binom{n_{(2)} - 1}{\tilde{m}_2} \left[q_{(22),n}\right]^{\tilde{m}_2} \left[1 - q_{(22),n}\right]^{n_{(2)} - \tilde{m}_2 - 1} \\ \to & \rho_{(21)}^{\tilde{m}_1 - 1} \frac{\exp\{-\rho_{(21)}\}}{(\tilde{m}_1 - 1)!} \rho_{(22)}^{\tilde{m}_2} \frac{\exp\{-\rho_{(22)}\}}{\tilde{m}_2!}, \end{aligned}$$

where the second equality follows from conditional independence in link formation under Condition R2, and the convergence is due to Poisson approximation of a binomial distribution. Similar derivation for the other case with $\tilde{m} = (\bar{n}_1, \bar{n}_2)$ implies similar results, only with probability mass functions replaced by survival functions in the limit. Thus $q^*(m'|x', x, m)$ exists for all $x, x' \in X$ and $m, m' \in M$. Note that in this example, $q^*(.|x', x, m)$ depends on x', x but not m, which is consistent with the rank condition for identification presented in Section 3.2.

We now show that asymptotic uncorrelation conditions in Assumption 6 hold in this model. Let $\xi_n \equiv (n_{(1)}, n_{(2)})$, where $n_{(k)} \equiv \#\{i \in N : x_i = k\}$. For a given sample size n and a pair of fixed values (x, m), the law of total covariance implies:

$$C_n(\iota_i(x, m), \iota_j(x, m)) = E_n\left[C_n(\iota_i(x, m), \iota_j(x, m) | X_i, X_j, \xi_n)\right] + C_n(\mu_{i,n}, \mu_{j,n}), \tag{22}$$

where $\mu_{i,n} \equiv E_n[\iota_i(x, m) | X_i, X_j, \xi_n]$ and likewise for $\mu_{j,n}$.

First, note that for $k, \ell \in \{1,2\}$ and $\bar{m}, \tilde{m} \in M$, and conditional on ξ_n (which we suppress in notation for now),

$$\begin{aligned} & P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l\} \\ = \;& P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 1\} q_{(k\ell),n} \\ + \;& P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\}(1 - q_{(k\ell),n}). \end{aligned}$$

Because $q_{(k\ell),n} \to 0$ as $n \to \infty$, the difference between the right-hand above and the sequence $P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\}$ diminishes to zero as $n \to \infty$. For example, for $k \neq l$ and any uncensored values of $\bar{m}, \tilde{m}$ in M,

$$\begin{aligned} & P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\} \\ = & \binom{n_{(k)} - 1}{\bar{m}_k} \left[q_{(kk),n}\right]^{\bar{m}_k} \left[1 - q_{(kk),n}\right]^{n_{(k)} - \bar{m}_k - 1} \binom{n_{(\ell)} - 1}{\bar{m}_\ell} \left[q_{(k\ell),n}\right]^{\bar{m}_\ell} \left[1 - q_{(k\ell),n}\right]^{n_{(\ell)} - \bar{m}_\ell - 1} \\ & \binom{n_{(k)} - 1}{\tilde{m}_k} \left[q_{(lk),n}\right]^{\tilde{m}_k} \left[1 - q_{(lk),n}\right]^{n_{(k)} - \tilde{m}_k - 1} \binom{n_{(\ell)} - 1}{\tilde{m}_\ell} \left[q_{(\ell\ell),n}\right]^{\tilde{m}_\ell} \left[1 - q_{(\ell\ell),n}\right]^{n_{(\ell)} - \tilde{m}_\ell - 1}. \end{aligned}$$

Again, by an application of the Poisson approximation of a Binomial distribution and the Le Cam's Theorem, we have:

$$P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\} - P_n\{m_i = \bar{m} | x_i = k\} P_n\{m_j = \tilde{m} | x_j = l\} \to 0$$

as $n \to \infty$ for all $\bar{m}, \tilde{m} \in M$. This implies the first term on the right-hand side of (22) converges to zero as the network size increases.

Next, note the second term on the right-hand side of (22) is:

$$C_n(\mu_{i,n}, \mu_{j,n}) = E_n\left(\mu_{i,n}, \mu_{j,n}\right) - E_n(\mu_{i,n}) E_n(\mu_{j,n}). \tag{23}$$

In what follows, we sometimes suppress the dependence of ι_i, ι_j on the fixed values (x, m) to simplify notation when there is no ambiguity. By the law of iterated expectation,

$$\begin{aligned} E_n\left(\mu_{i,n}, \mu_{j,n}\right) &= \sum\nolimits_{x_i} \sum\nolimits_{x_j} \Pr\{(X_i, X_j) = (x_i, x_j)\} E_n\left[\mu_{i,n}, \mu_{j,n} \middle| (X_i, X_j) = (x_i, x_j)\right] \\ &= \Pr\{(X_i, X_j) = (x, x)\} E_n\left[\mu_{i,n}, \mu_{j,n} \middle| (X_i, X_j) = (x, x)\right], \end{aligned} \tag{24}$$

where the last equality holds because $E_n\left[\mu_{i,n}\mu_{j,n} \middle| (X_i, X_j) = (x_i, x_j)\right] = 0$ if either $x_i \neq x$ or $x_j \neq x$.[13] Furthermore, independent dyadic link formation in (R2) implies:

$$\mu_{i,n}(x, m) = \sum\nolimits_{d \in \{0,1\}} E_n[\iota_i(x, m) | X_i, X_j, \xi_n, g_{ij} = d] P_n(g_{ij} = d | X_i, X_j).$$

[13]Recall $\mu_{i,n}, \mu_{j,n}$ are conditional expectations of two indicator functions evaluated at fixed values (x, m), which we suppress in notation.

A similar decomposition holds for $\mu_{j,n}$, with ι_i replaced by ι_j above.

Without loss of generality, consider the case with $x = 1$. Under (R1) and (R2),

$$\begin{aligned} & E_n[\iota_i(1,m)|(X_i,X_j) = (1,1), \xi_n = (n_{(1)}, n_{(2)}), g_{ij} = 0] \\ = \; & E_n\left[1\{M_i = (m_1,m_2)\}|(X_i,X_j) = (1,1), \xi_n = (n_{(1)}, n_{(2)}), g_{ij} = 0\right] \\ = \; & \binom{n_{(1)}-2}{m_1}\left[q_{(11),n}\right]^{m_1}\left[1-q_{(11),n}\right]^{n_{(1)}-m_1-2}\binom{n_{(2)}}{m_2}\left[q_{(12),n}\right]^{m_2}\left[1-q_{(12),n}\right]^{n_{(2)}-m_2} \end{aligned} \tag{25}$$

which, as $n_{(1)}, n_{(2)} \to \infty$, converges to:

$$\rho_{(11)}^{m_1}\frac{\exp\{-\rho_{(11)}\}}{(m_1)!}\rho_{(12)}^{m_2}\frac{\exp\{-\rho_{(12)}\}}{(m_2)!} \equiv \delta_1,$$

according to the Le Cam Theorem.

By a similar argument, $E_n\left[1\{M_i = (m_1,m_2)\}|(X_i,X_j) = (1,1), \xi_n = (n_{(1)}, n_{(2)}), g_{ij} = 1\right]$ converges to another finite constant similar to δ_1 as $n_{(1)}, n_{(2)} \to \infty$, only with m_1 replaced by $m_1 - 1$. Because the link formation rates $q_{(kl),n} \to 0$ for all $k,l \in \{1,2\}$, we know $\mu_{i,n}(1,m)$ converges to δ_1 at $(X_i,X_j) = (1,1)$ as the network grows. By the same argument, $\mu_{j,n}(1,m)$ converges to the *same* constant δ_1 at $(X_i,X_j) = (1,1)$ as the network grows.

With X_i i.i.d. across i and $\Pr\{X_i = k\} = p_{(k)}$ for $k = 1,2$ fixed at all sample sizes, we have $\mu_{i,n}(1,m) \xrightarrow{p} \delta_1$ and $\mu_{j,n}(1,m) \xrightarrow{p} \delta_1$ for $(X_i,X_j) = (1,1)$ as the network grows $n \to \infty$. (See a formal proof of the claim below.) An analogous argument shows $\mu_{i,n}(2,m) \xrightarrow{p} \delta_2$ and $\mu_{j,n}(2,m) \xrightarrow{p} \delta_2$ for $(X_i,X_j) = (2,2)$ as $n \to \infty$, where δ_2 is similar to δ_1, only with $\rho_{(1k)}$ replaced by $\rho_{(2k)}$. It then follows from (24) that, for any (x,m),

$$E_n[\mu_{i,n}(x,m)\mu_{j,n}(x,m)] \to \left[p_{(x)}\right]^2(\delta_x)^2 \text{ as } n \to \infty. \tag{26}$$

Next, we show convergence of $E_n(\mu_{i,n})$ and $E_n(\mu_{j,n})$. By the law of iterated expectation,

$$\begin{aligned} E_n\left[\mu_{i,n}(x,m)\right] &= \sum\nolimits_{(x_i,x_j)} \Pr\{(X_i,X_j) = (x_i,x_j)\}E_n\left[\mu_{i,n}(x,m)|\,(X_i,X_j) = (x_i,x_j)\right] \\ &= \Pr\{X_i = x\}\left\{\sum_{x_j}\Pr\{X_j = x_j\}E_n\left[\mu_{i,n}(x,m)|\,(X_i,X_j) = (x,x_j)\right]\right\} \end{aligned} \tag{27}$$

W.L.O.G., consider the case with $(x,m) = (1,m)$. We already showed $\mu_{i,n}(1,m) \xrightarrow{p} \delta_1$ for

$(X_i, X_j) = (1,1)$ as $n \to \infty$, so it remains to consider the case with $(X_i, X_j) = (1,2)$. Note:

$$\begin{aligned} & E_n[\iota_i(1,m)|(X_i,X_j) = (1,2), \xi_n = (n_{(1)}, n_{(2)}), g_{ij} = 0] \\ = \; & E_n\left[1\{M_i = (m_1, m_2)\}|(X_i,X_j) = (1,2), \xi_n = (n_{(1)}, n_{(2)}), g_{ij} = 0\right] \\ = \; & \binom{n_{(1)}-1}{m_1} q_{(11),n}^{m_1}\left[1 - q_{(11),n}\right]^{n_{(1)}-m_1-1} \binom{n_{(2)}-1}{m_2} q_{(12),n}^{m_2}\left[1 - q_{(12),n}\right]^{n_{(2)}-m_2-1}, \end{aligned}$$

which converges to the *same* constant δ_1 as in the case of $(X_i, X_j) = (1,1)$ in (25). (This is because the growing sequence $n_{(1)}, n_{(2)}$ has no role in the limit, as long as $n_{(k)} \times q_{(kl),n}$ converges to $\rho_{(kl)}$ as the network grows.) With diminishing dyadic link formation rates, $P_n\{g_{ij} = 0|X_i, X_j\}$ converges to one as the network grows. As before, by the law of iterated expectation, we have $\mu_{i,n}(1,m) \xrightarrow{p} \delta_1$ for $(X_i, X_j) = (1,2)$. Most importantly, this probability limit is *the same* as for $(X_i, X_j) = (1,1)$. This implies $E_n\left[\mu_{i,n}(x,m)|(X_i,X_j) = (x, x_j)\right]$ converges to the same constant δ_x, *regardless of* the value of x_j.

It follows from (27) that $E_n\left[\mu_{i,n}(x,m)\right] \to p_{(x)}\delta_x$ for any fixed value of x as $n \to \infty$. By a symmetric argument, $E_n\left[\mu_{j,n}(x,m)\right] \to p_{(x)}\delta_x$. Consequently,

$$E_n\left[\mu_{i,n}(x,m)\right] E_n\left[\mu_{j,n}(x,m)\right] \to \left[p_{(x)}\right]^2 (\delta_x)^2 \text{ as } n \to \infty. \tag{28}$$

Together, (23), (26) and (28) imply the second term on the right-hand side of (22) diminishes:

$$C_n(\mu_{i,n}, \mu_{j,n}) \to 0 \text{ as } n \to \infty.$$

Finally, the condition that

$$\mathbf{C}_n(\iota_j(x', m')\iota_i(x,m)g_{ij}, \iota_\ell(x', m')\iota_t(x,m)g_{lt}) \to 0 \; \forall x, x' \in X, m, m' \in M$$

as $n \to \infty$ for $\{i, j\} \cap \{t, l\} = \emptyset$ in this context follows from similar derivation, which we omit in the text for brevity.

Some intermediate results above make use of the following claim, which we now state and prove formally. Recall that $n \equiv \#N$ denotes the network (sample) size, and $\{X_i\}_{i \le n}$ are i.i.d. with $\Pr\{X_i = k\} = p_{(k)}$ for $k = 1, 2$, where $p_{(k)}$ are fixed constants invariant to the sample size. Also recall that $\xi_n \equiv (\xi_{n,1}, \xi_{n,2})$, where $\xi_{n,k} \equiv \#\{i \in N : X_i = k\}$ for $k = 1, 2$. Let $\mathcal{N}_n = \{(n_1, n_2) \in \mathbb{N}_+^2 : n_1 + n_2 = n\}$ denote the support of ξ_n.

Claim. Let $f_n : \mathcal{N}_n \to \mathbb{R}$ be a sequence of real-valued functions with domain $\mathcal{N}_n$. If $f_n(n_1, n_2)$ converges to some constant $c < \infty$ as $n_1, n_2 \to \infty$, then $f_n(\xi_n) \xrightarrow{p} c$ as $n \to \infty$.

Proof. By assumption, for any $\varepsilon > 0$, there exist integers $\bar{n}_1, \bar{n}_2 < \infty$ so that $|f_n(n_1, n_2) - c| \le \varepsilon$ for all $n_1 > \bar{n}_1$ and $n_2 > \bar{n}_2$. Hence

$$\Pr\{\xi_{n,1} > \bar{n}_1 \text{ and } \xi_{n,2} > \bar{n}_2\} \le \Pr\{|f_n(\xi_n) - c| \le \varepsilon\},$$

or, equivalently,

$$\begin{aligned}\Pr\{|f_n(\xi_n) - c| > \varepsilon\} \le & \quad \Pr\{\xi_{n,1} \le \bar{n}_1 \text{ or } \xi_{n,2} \le \bar{n}_2\} \\ \le & \quad \Pr\{\xi_{n,1} \le \bar{n}_1\} + \Pr\{\xi_{n,2} \le \bar{n}_2\},\end{aligned}$$

where $\Pr\{\xi_{n,1} \le \bar{n}_1\} = \sum_{s=0}^{\bar{n}_1} \phi_{s,n}$ with $\phi_{s,n} \equiv \binom{n}{s} p_{(1)}^s [1 - p_{(1)}]^{n-s}$ and likewise for $\Pr\{\xi_{n,2} \le \bar{n}_2\}$. For $s = 0$, $\phi_{s,n} = [1 - p_{(1)}]^n \to 0$ as $n \to \infty$. For any $1 \le s \le \bar{n}_1$, $\binom{n}{s}$ is bounded above by $n^s / (s!)$. Hence $\phi_{s,n} \le \overline{C} n^s [1 - p_{(1)}]^n$, with $\overline{C} \equiv \frac{p_{(1)}^s}{(s!)[1-p_{(1)}]^s}$ being a finite constant, and

$$\log \phi_{s,n} \le \log \overline{C} + s \log(n) + n \log \left[1 - p_{(1)}\right].$$

For any finite s, the right-hand side diverges to $-\infty$ as $n \to \infty$ (because $n^{-1} \log n \to 0$ as $n \to \infty$). Therefore, for each $0 \le s \le \bar{n}_1$, we have $\phi_{s,n} \to 0$ as $n \to \infty$. This implies $\Pr\{\xi_{n,1} \le \bar{n}_1\} \to 0$ as $n \to \infty$. By the same argument, $\Pr\{\xi_{n,2} \le \bar{n}_2\} \to 0$ as $n \to \infty$. Hence for any $\varepsilon > 0$,

$$\Pr\{|f_n(\xi_n) - c| > \varepsilon\} \to 0 \text{ as } n \to \infty.$$

That is, $f_n(\xi_n) \xrightarrow{p} c$ as $n \to \infty$. $\square$

C. Asymptotic Distribution of the Two-Step Estimator

In this part of the appendix, we sketch a heuristic discussion about the asymptotic distribution of our two-step m-estimator. As noted in the text, our goal here is not to show how to conduct formal inference. Instead, we intend to provide a framework for understanding and deriving the estimator's asymptotic properties. Full characterization of its limiting variance under primitive conditions, as well as the construction of standard errors, would require more elaborate conditions that restrict the form and magnitude of dependence between individual-specific variables (i.e., actions and characteristics) over the network. We leave these tasks for future research.

Let $\hat{\rho}$ denote the vector of all first-stage estimates in $\hat{\lambda}_i$ and $\hat{\chi}_{\ell,i}$. That is,

$$\hat{\rho} = \begin{pmatrix} n^{-1}\sum_i \iota_i(x,m) \\ n^{-1}\sum_i a_i\iota_i(x,m) \\ \left(\frac{1}{n(n-1)}\sum_{j\neq i}\iota_j(x',m')\iota_i(x,m)g_{ij}\right)_{x',m'} \\ \left(\frac{1}{n(n-1)}\sum_{j\neq i}1\{x_j=x'\}\iota_i(x,m)g_{ij}\right)_{x'} \end{pmatrix}_{x,m}.$$

Let $\nabla_\delta \hat{G}_n(\delta) = n^{-1}\sum_i \Gamma_i(\delta;\hat{\rho})$, with

$$\Gamma_i(\delta;\hat{\rho}) = 2\bar{q}_i(\delta;\hat{\rho})[\eta_i(\theta), \hat{\mathbf{c}}_i, \gamma\nabla_\theta\eta_i(\theta)]$$

where

$$\bar{q}_i(\delta;\hat{\rho}) = \hat{\lambda}_i - \gamma\eta_i(\theta) - \sum_k \mathbf{1}_{k,i}\sum_\ell \hat{\chi}_{\ell,i}\beta_{k\ell}$$

and $\hat{\mathbf{c}}_i$ is such that $2\bar{q}_i(\delta;\hat{\rho})\hat{\mathbf{c}}_i$ is the deriative of the objective function with respect to the vector of coefficients β. By the first-order condition and a mean-value expansition,

$$\sqrt{n}\nabla_\delta\hat{G}_n(\delta) + \nabla^2_{\delta,\delta}\hat{G}_n(\tilde{\delta})\sqrt{n}(\hat{\delta}-\delta) = o_p(1),$$

where $\tilde{\delta}$ is some intermediate value between $\hat{\delta}$ and δ. Assume:

D1. $\sup_{x\in X,m\in M}\|\hat{\rho}-\rho\| = o_p(n^{-1/4})$.

D2. $n^{-1}\sum_i \nabla_\delta\Gamma_i(\delta;\rho) - \Psi_n = o_p(1)$, where $\Psi_n \equiv \mathbf{E}_n[\frac{1}{n}\sum_i \nabla_\delta\Gamma_i(\delta;\rho)]$ is full-rank for all n.

D3. $\nabla^2_{\rho,\rho}\Gamma_i(\delta;\rho)$ exists and is bounded over an open neighborhood around ρ almost surely under p^*.

D4. $n^{-1}\sum_i \nabla_\rho\Gamma_i(\delta;\rho) - \Phi_n = O_p(n^{-1/4})$, where $\Phi_n \equiv \mathbf{E}_n[\frac{1}{n}\sum_i \nabla_\rho\Gamma_i(\delta;\rho)]$ has full-rank for each n.

D5. $\hat{\rho}-\rho = n^{-1}\sum_i \psi_{n,i} + o_p(n^{-1/2})$ for some $\psi_{n,i}$ determined by (δ,ρ) such that $\mathbf{E}_n(\psi_{n,i}) = 0$ for all n.

D6. $\Lambda_n^{-1}\left\{\frac{1}{\sqrt{n}}\sum_i[\Gamma_i(\delta;\rho) + \Phi_n \times \psi_{n,i}]\right\} \xrightarrow{d} N(0,I)$, where I is the identity matrix and Λ_n a sequence of positive semi-definite matrices.

The scaling matrix Λ_n in D6 captures dependence of Γ_i across individual i over the network. Characterization of this matrix requires invoking primitive conditions to establish a notion of weak dependence between observed individual outcomes over the network. This would allow us to apply some central limit theorem for weak dependence to the relevant sample averages as the sample (network) size increases. These entail more elaborate work, and are suitable for further investigation in another paper.

Lemma C1. Under D1 and D2, $\nabla^2_{\delta,\delta}\hat{G}_n(\tilde{\delta}) - \Psi_n \xrightarrow{p} 0$ whenever $\tilde{\delta} \xrightarrow{p} \delta$.

Proof of Lemma C1. Let $\Gamma_i(\delta, \rho)$ be defined in a way that is similar to $\Gamma_i(\delta; \hat{\rho})$, only with the first-stage estimates $\hat{\rho}$ replaced by the probability limit ρ. That is,

$$\begin{aligned}\Gamma_i(\delta;\rho) \quad &= \quad 2\left(\lambda^*(x_i,m_i) - \gamma\eta(x_i,m_i;\theta) - \sum\nolimits_k \mathbf{1}_{k,i}\sum\nolimits_\ell \chi^*_\ell(x_i,m_i)\beta_{k\ell}\right) \\ &\quad\quad \times[\eta(x_i,m_i;\theta), \mathbf{c}_i, \gamma\nabla_\theta\eta(x_i,m_i;\theta)].\end{aligned}$$

By definition,

$$\nabla^2_{\delta,\delta}\hat{G}_n(\tilde{\delta}) - \Psi_n = \underbrace{n^{-1}\sum\nolimits_i\left[\nabla_\delta\Gamma_i(\tilde{\delta};\hat{\rho}) - \nabla_\delta\Gamma_i(\delta;\rho)\right]}_{A} + \underbrace{n^{-1}\sum\nolimits_i \nabla_\delta\Gamma_i(\delta;\rho) - \Psi_n}_{B}.$$

The absolute value of the first term A on the right-hand side is bounded above by

$$\sup\nolimits_{x,m}|\nabla_\delta\Gamma(x,m;\tilde{\delta};\hat{\rho}) - \nabla_\delta\Gamma(x,m;\delta;\rho)|.$$

Note that the consistency of $\hat{\delta}$ implies $\tilde{\delta} \xrightarrow{p} \delta$. Hence it is bounded above by a term that is $o_p(1)$ because $\sup_{x\in X, m\in M}\|\hat{\rho} - \rho\| \xrightarrow{p} 0$. The second term B is $o_p(1)$ under condition D2. □

Lemma C2. Under D1, D3, D4 and D5,

$$\sqrt{n}\nabla_\delta\hat{G}_n(\delta) = n^{-1}\sum\nolimits_i[\Gamma_i(\rho) + \Phi_n \times \psi_{n,i}] + o_p(n^{-1/2}).$$

Proof of Lemma C2. Let $\Gamma_i(\hat{\rho})$ be shorthand for $\Gamma(x_i, m_i; \delta; \hat{\rho})$. By a second-order Taylor expansion, we can write

$$\begin{aligned}\nabla_\delta\hat{G}_n(\delta) \quad &= \quad n^{-1}\sum\nolimits_i\Gamma_i(\hat{\rho}) \\ &= \quad n^{-1}\sum\nolimits_i\left[\Gamma_i(\rho) + \nabla_\rho\Gamma_i(\rho)(\hat{\rho}-\rho) + \tfrac{1}{2}(\hat{\rho}-\rho)'\nabla^2_{\rho,\rho}\Gamma_i(\tilde{\rho})(\hat{\rho}-\rho)\right]\end{aligned}$$

for some $\tilde{\rho}$ between $\hat{\rho}$ and ρ. Under D3, the absolute value of $n^{-1}\sum_i(\hat{\rho}-\rho)'\nabla^2_{\rho,\rho}\Gamma_i(\tilde{\rho})(\hat{\rho}-\rho)$ on the right-hand side is bounded above by the product of a constant and $\sup_{x\in X, m\in M}\|\hat{\rho}-\rho\|^2$, where $\|.\|$ denotes the sup-norm. Under D1, this upper bound is $o_p(n^{-1/2})$. Under D4, $[n^{-1}\sum_i \nabla_\rho \Gamma_i(\rho) - \Phi_n] \times (\hat{\rho} - \rho) = o_p(n^{-1/2})$. The claim in the lemma follows from D5. $\square$

It then follows that $(\Psi_n^{-1}\Lambda_n\Psi_n^{-1})^{-1/2}\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{d} N(0, I)$ under D1-D5.