A Structural Econometric Analysis of Network Formation Games*

Shuyang Sheng†

October 2, 2016

Abstract

The objective of this paper is to identify and estimate network formation models using observed data on network structure. We characterize network formation as a simultaneous-move game, where the utility from forming a link depends on the structure of the network, thereby generating strategic interactions between links. Because a unique equilibrium may not exist, the parameters are not necessarily point identified. We leave the equilibrium selection unrestricted and propose a partial identification approach. We derive bounds on the probability of observing a subnetwork, where a subnetwork is the restriction of a network to a subset of the individuals. Unlike the standard bounds as in Ciliberto and Tamer (2009), these subnetwork bounds are computationally tractable in large networks provided we consider small subnetworks. The information in these bounds also converges as the network size approaches infinity. We provide Monte Carlo evidence that bounds from small subnetworks are informative in large networks.

JEL Classifications: C13, C31, C57, D85

KEYWORDS: Network formation, simultaneous-move games, multiple equilibria, subnetworks, partial identification, moment inequalities, simulation.

---

*This paper is a revision of Chapter 2 of my dissertation. I am very grateful to my advisor Geert Ridder for his enormously valuable advice and guidance. I also thank Aureo de Paula, Bryan Graham, Jinyong Hahn, Matthew Jackson, Rosa Matzkin, Roger Moon, Hashem Pesaran, Matthew Shum, Martin Weidner, Simon Wilkie, seminar participants at USC, UCLA, UCSD, JHU, Pittsburgh, Tilburg, CORE, California Econometrics Conference, NASM, NAWM, EMES for helpful discussions and comments. Financial support from the USC Graduate School Dissertation Completion Fellowship is acknowledged. All the errors are mine.

†Department of Economics, UCLA, Los Angeles, CA 90095. E-mail: ssheng@econ.ucla.edu.

1 Introduction

Social and economic networks influence a variety of individual behaviors and outcomes, including educational achievement (Calvó-Armengol, Patacchini, and Zenou (2009)), employment (Calvó-Armengol and Jackson (2004)), technology adoption (Conley and Udry (2010)), consumption (Moretti (2011)), and smoking (Nakajima (2007)). As networks are often the result of individual decisions, understanding the formation of networks is important for the investigation of network effects. Despite that the theoretical literature on network formation having flourished in the past decades (see Jackson (2008) and Goyal (2007) for a survey), econometric studies on the identification and estimation of network formation models are still at an infant stage. The objective of this paper is to provide insight into this latter area. More precisely, assume that we observe the network structure, i.e., who is linked with whom. We propose new methods to identify and estimate the structural parameters in the model of network formation.

The statistical analysis of network formation dates back to the seminal work of Erdős and Rényi (1959), who proposed a random graph model where links are formed independently with a fixed probability. Statisticians later extended the Erdős-Rényi model to allow for dependence between links and developed a large class of exponential random graph models (ERGM) (e.g., Snijder (2002)). While ERGMs may well fit the observed network statistics, they usually lack microfoundations which are essential for counterfactual analysis. Alternatively, economists view network formation as the optimal choices of individuals that maximize their utilities. A simple and widely used empirical approach in this spirit is to employ a dyadic regression, where the formation of a link is modeled as a binary choice of the pair involved (e.g., Fafchamps and Gubert (2007), Mayer and Puller (2008)). In order to treat links in a network as independent observations, this approach needs to assume that there is no spillover from indirect friends (e.g., friends of friends), which could be restrictive in many applications given the prevalence of clustering (e.g., Jackson and Rogers (2007), Jackson, Barraquer and Tan (2012)). Graham (2016) extends dyadic regressions by allowing for individual fixed effects which can create interdependence between links. A more general class of network formation models permits utility externalities from indirect friends, thereby giving rise to strategic interactions between links (Christakis, Fowler, Imbens, and Kalyanaraman (2010), Mele (2011), Boucher and Mourifié (2013), Miyauchi (2013), Leung (2015), Ridder and Sheng (2016), De Paula, Richards-Shubik and Tamer (2015), Menzel (2016b)). A contribution of this paper is to provide a different approach for the identification and estimation of such strategic network formation models.

A crucial problem in the identification of network formation models with strategic interactions is the presence of multiple equilibria. Bouncher and Mourifié (2013) get around this problem by assuming there is a unique equilibrium in the observed data. Christakis et al. (2010) and Mele (2011) circumvent the multiplicity issue by considering a sequential model where each link is formed in a random sequence and myopically. The Markov chain of networks achieved in each period may converge to a unique stationary distribution over the collection of equilibrium networks. Employing the stationary distribution to construct the data likelihood is then equivalent to imposing implicitly an equilibrium selection mechanism in the corresponding static model (Young (1993), Jackson and Watts (2002)). Unlike these studies, we admit multiple equilibria and do not impose restrictive assumptions on equilibrium selection. Since a unique equilibrium may not exist in our setting, the parameters are not necessarily point identified. We propose a partial identification approach and examine what we can learn about the parameters from bounds on conditional choice probabilities. The study closest to ours is by Miyauchi (2013), who considers partial identification as well. Miyauchi derives his bounds from a partial ordering of equilibrium networks under a nonnegative externality assumption, while our bounds hold for more general utility functions.

The estimation of network formation models is computationally challenging because the number of possible networks is enormous: for n individuals the number of possible undirected networks is $2^{n(n-1)/2}$. In ERGMs, parameter estimation relies crucially on sampling networks from exponential family distributions. Given the huge space of possible networks, the sampling is typically carried out using Markov Chain Monte Carlo (MCMC) methods. However, the mixing time of MCMC is $O(e^n)$ unless links are approximately independent, in which case the model is not appreciably different from the Erdős-Rényi model (Bhamidi, Bresler, and Sly (2011)). Chandrasekhar and Jackson (2013) provide Monte Carlo evidence that slow convergence of MCMC leads to poor performance of ERGMs. In sequential models of network formation, likelihoods constructed using stationary distributions may be computationally intractable because such likelihoods typically include a sum over all possible networks (e.g., Mele (2011)). While MCMC methods can be used to avoid computing intractable likelihoods, they need to simulate networks from the stationary distributions where the mixing rate can be as slow as $O(e^n)$. Hence, sequential models suffer from the same computational problem as in ERGMs.

In our model, the computation of the bounds may be intractable as well because it requires checking equilibrium conditions for all possible network configurations. We propose a completely new approach to tackle this computational problem. The idea is to make use of subnetworks. A subnetwork is the restriction of a network to a subset of the individuals. Under the equilibrium concept we consider (i.e., pairwise stability proposed by Jackson and Wolinsky (1996)), we can derive the best possible bounds on the probability of observing a subnetwork. Under our utility specification these subnetwork bounds are computationally tractable even in large networks as long as we only consider small subnetworks. This approach only needs choice probabilities within subnetworks, so it is still applicable if we do not observe an entire network, but links in subnetworks.

The subnetwork bounds remain useful as networks grow in size. Under assumptions that ensure exchangeability in observed networks, inequalities from subnetworks of any size converge as n tends to infinity. Therefore, bounds from small subnetworks remain informative about the parameters in large networks. It is worth pointing out that our approach differs substantially from a recent strand of literature on large networks, which typically assumes that a single large network is observed (Leung (2015), Ridder and Sheng (2016), De Paula, Richard-Shubik and Tamer (2015), Menzel (2016b)). By assuming many networks, our approach does not need the restrictions that these studies may have to impose to control for the dependence between links and can be seen as complementary to these studies.

The estimation and inference of the identified set defined by the subnetwork inequalities is a straightforward application of the literature on partially identified models (e.g., Chernozhukov, Hong and Tamer (2007), Andrews and Soares (2010), Romano and Shaikh (2010), Andrews and Jia (2012)). Exchangeability implies that subnetworks in a network of the same size follow the same distribution, so the subnetwork choice probabilities in the moment inequalities can be estimated using randomly selected subnetworks of a given size. The bounds do not have a closed form. We propose how to compute them by simulation.

Other Related Literature Our paper is related to the econometric literature on static games of complete information (e.g., Bresnahan and Reiss (1991), Tamer (2003), Ciliberto and Tamer (2009), Bajari, Hong, and Ryan (2010), Bajari, Hahn, Hong, and Ridder (2011)). Such games often face the identification problem due to the prevalence of multiple equilibria. To avoid imposing restrictions on equilibrium selection, econometricians have applied partial identification to such games (e.g., Andrews, Berry and Jia (2004), Pakes, Porter, Ho and Ishii (2006), Berry and Tamer (2006), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011)). However, most studies look at simple entry games where the number of agents is small. We contribute to this literature by developing a partial identification approach to network formation games where the number of agents can be large, so standard probability bounds are computationally intractable. By focusing on bounds from small subnetworks, we can achieve computational feasibility. This idea may shed light on the analysis of other games with a large number of agents (e.g., matching games) and provide a new perspective on reducing the dimensionality of those models. Related literature includes Menzel (2015, 2016a).

The remainder of the paper is organized as follows. Section 2 develops the model. Section 3 addresses the multiple equilibrium problem and proposes the partial identification approach. Section 4 develops the subnetwork approach. We derive the subnetwork inequalities in Section 4.1 and analyze their asymptotic properties in Section 4.2. Section 5 discusses the estimation methods. Section 6 discusses how to compute the bounds. Section 7 conducts a Monte Carlo study, and Section 8 concludes the paper.

2 A Model of Network Formation

In this section, we develop the network formation model. Let $[n] = \{1, 2, ..., n\}$ be the set of individuals who can form links. The links are undirected in the sense that forming a link requires the consent of both individuals involved in the link, but severing a link can be unilateral. This is the natural setting in the context of friendship networks, and for that reason we call linked individuals friends.

The links form a network, which we denote by $G \in \mathcal{G}$. It is an $n \times n$ binary matrix, where $G_{ij} = 1$ if i and j are friends, and 0 otherwise for all $i \neq j$. Since we consider undirected links, G is a symmetric matrix. We normalize $G_{ii} = 0$ for all i.

Utility Each individual i has a $d_x \times 1$ vector of observed attributes X_i (e.g., gender, age, race) and an $(n-1) \times 1$ vector of unobserved (to researchers) preferences $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{i,i-1}, \varepsilon_{i,i+1}, \ldots, \varepsilon_{in})'$, where ε_{ij} is i's preference for link ij. Let $X = (X_1', \ldots, X_n')'$ and $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_n')'$. The utility of individual i in a network in general depends on the network configuration G, the observed attributes X, and i's unobserved preferences ε_i, i.e.,

$$U_i(G, X, \varepsilon_i).$$

For any $i \neq j$, we decompose G into (G_{ij}, G_{-ij}), where $G_{-ij} \in \mathcal{G}_{-ij}$ is the network obtained from G by removing link ij. Then the marginal utility of i from forming a link with j is

$$\Delta U_{ij}(G_{-ij}, X, \varepsilon_i) = U_i(1, G_{-ij}, X, \varepsilon_i) - U_i(0, G_{-ij}, X, \varepsilon_i). \tag{1}$$

In this paper, we consider the utility specification

$$\begin{aligned} U_i(G, X, \varepsilon_i) &= \sum_{j=1}^{n} G_{ij}\left(u(X_i, X_j; \beta) + \varepsilon_{ij}\right) + \frac{1}{n-2}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ k\neq i}}^{n} G_{ij}G_{jk}\gamma_1 \\ &\quad + \frac{1}{n-2}\sum_{j=1}^{n}\sum_{k=j+1}^{n} G_{ij}G_{ik}G_{jk}\gamma_2, \end{aligned} \tag{2}$$

where $u(X_i, X_j; \beta) = \beta_0 + \beta_1' X_i + \beta_2' |X_i - X_j|$. In this specification, the first term is the utility (net cost) from direct friends, where the term $|X_i - X_j|$ is to capture the homophily effect, which says that people tend to make friends with those who are similar to them (Currarini, Jackson and Pin (2009), Christakis et al. (2010)). In addition to the direct-friend effects, (2) also allows for the effects of indirect friends. The second term in (2) captures the utility from i's friends of friends, and the third term captures the additional utility if i and i's friend have friends in common,[1] where γ_1 and γ_2 are constants in $\mathbb{R}$. Hence, if we consider the marginal utility of i from forming a link with j, which is given by

$$\Delta U_{ij}(G_{-ij}, X_i, X_j, \varepsilon_{ij}) = u(X_i, X_j; \beta) + \frac{1}{n-2}\sum_{\substack{k=1 \\ k\neq i,j}}^{n} G_{jk}\gamma_1 + \frac{1}{n-2}\sum_{\substack{k=1 \\ k\neq i,j}}^{n} G_{ik}G_{jk}\gamma_2 + \varepsilon_{ij}, \tag{3}$$

[1] The latter is motivated by the clustering hypothesis, which says that if two individuals have friends in common, they are more likely to be friends than if links are formed randomly (Jackson and Rogers (2007), Jackson (2008), Christakis et al. (2010), Jackson et al. (2012)).

then it consists of not only the direct utility from j, but also the indirect utility from j's other friends and i, j's friends in common. This utility function follows closely the specification in Christakis et al. (2010).[2] It is also related to the specifications in Mele (2011) and Goyal and Joshi (2006), but is more general than both.[3] In addition, note that the effects of friends of friends and friends in common are normalized by $n-2$. We show in Section 4.2 that under the normalization both sum terms in (3) converge as $n \to \infty$ so these effects remain stable in large networks.[4]

Equilibrium Given the utilities, individuals choose friends simultaneously as in the link-announcement game (Myerson (1991), Jackson (2008)). We assume that individuals observe X and ε, so it is a complete-information game. Depending whether transfers are allowed for, each individual announces a set of *intended links* or *intended transfers*. Under nontransferable utility (NTU), a link is formed if both individuals intend to form it, while under transferable utility (TU) a link is formed if the sum of the two transfers for it is nonnegative.

The equilibrium concept we consider in the paper is *pairwise stability* (Jackson and Wolinsky (1996) for NTU, Bloch and Jackson (2006, 2007) for TU). We say a network is pairwise stable if no pair of individuals wants to create a new link, and no individual wants to sever an existing link. Formally,

Definition 2.1 *A network G is pairwise stable (PS) under NTU if*

1. *for any $G_{ij} = 1$, $\Delta U_{ij}(G_{-ij}, X_i, X_j, \varepsilon_{ij}) \geq 0$ and $\Delta U_{ji}(G_{-ij}, X_j, X_i, \varepsilon_{ji}) \geq 0$;*

2. *for any $G_{ij} = 0$, $\Delta U_{ij}(G_{-ij}, X_i, X_j, \varepsilon_{ij}) > 0 \implies \Delta U_{ji}(G_{-ij}, X_j, X_i, \varepsilon_{ji}) < 0$.*

Definition 2.2 *A network G is pairwise stable (PS) under TU if*

1. *for any $G_{ij} = 1$, $\Delta U_{ij}(G_{-ij}, X_i, X_j, \varepsilon_{ij}) + \Delta U_{ji}(G_{-ij}, X_j, X_i, \varepsilon_{ji}) \geq 0$;*

2. *for any $G_{ij} = 0$, $\Delta U_{ij}(G_{-ij}, X_i, X_j, \varepsilon_{ij}) + \Delta U_{ji}(G_{-ij}, X_j, X_i, \varepsilon_{ji}) \leq 0$.*

[2] Christakis et al. (2010) allow for nonlinear effects from friends of friends and friends in common. Our specification is a linear version of theirs. However, with linearity we can establish the existence of equilibrium, which is an open question for the specification they use.

[3] Mele (2011) considers a linear utility function which does not allow for the effects of friends in common. Goyal and Joshi (2006) assumes that the direct-friend effects are homogeneous across individuals.

[4] A referee suggested normalizing these sum terms at the rate they converge. We are grateful to this insightful suggestion.

In the sequel we consider both NTU and TU and use the term "pairwise stability" to mean pairwise stability under NTU or TU, depending on the context.

Since we allow for utility interdependence, the pairwise stability condition leads to a simultaneous discrete choice model, i.e.,

$$G_{ij} = 1\left\{\Delta U_{ij}(G_{-ij}, X_i, X_j, \varepsilon_{ij}) \geq 0, \Delta U_{ji}\left(G_{-ij}, X_j, X_i, \varepsilon_{ji}\right) \geq 0\right\}, \ \forall i \neq j, \tag{4}$$

under NTU and

$$G_{ij} = 1\left\{\Delta U_{ij}(G_{-ij}, X_i, X_j, \varepsilon_{ij}) + \Delta U_{ji}(G_{-ij}, X_j, X_i, \varepsilon_{ji}) \geq 0\right\}, \ \forall i \neq j, \tag{5}$$

under TU,[5] where the choice of a link G_{ij} depends on the choices of others G_{-ij}. This indicates that we cannot treat each link as a single observation and use a dyadic regression because G_{-ij} is endogenous in the model, so can be correlated with $(\varepsilon_{ij}, \varepsilon_{ji})$. What further complicates the statistical inference of (4) and (5) is that there may be multiple equilibria, which will affect the identification of the parameters.

The existence of pairwise stable networks is also not guaranteed. According to Jackson and Watts (2002, Lemma 1), for any utility function there is either a PS network or a closed cycle.[6,7] In the appendix we give an example where there is no PS network, but a closed cycle. A closed cycle represents a situation in which *for the given utilities* individuals never reach a stable state and constantly switch between forming and severing links, which is unlikely to occur in real applications. To ensure that our model yields an appropriate solution, we need a utility function such that for any parameter value, X and ε, there exists a PS network.

Most results in the network literature on the existence of PS networks do not allow for heterogeneity among individuals and thus are unsuitable for our analysis.[8] Jackson and Watts (2001) and Hellmann (2012) provide general conditions under which a PS

[5]Equations (4) and (5) differ slightly from Definitions 2.1 and 2.2 in the indifference case, but the discrepency is negligible when ε follows a continuous distribution.

[6]A *closed cycle* is a collection of networks such that: (i) for any two networks in the collection there is an improving path from one to the other; and (ii) no improving path starting from a network in the collection leads to a network outside. Here an *improving path* is a sequence of networks in which two consecutive networks differ by one link, and adding (or deleting) the link in the succeeding network is beneficial for the individuals involved. See Jackson and Watts (2002) for rigorous definitions.

[7]The original result in Jackson and Watts (2002) was proved under NTU. It is easy to show that their result also holds under TU.

[8]See, for example, Belleflamme and Bloch (2004), Goyal and Joshi (2006).

network exists. We apply their conditions and provide existence results for the utility function in (2). The insight of these results is that (1) under TU the model permits a representation as a potential game (Monderer and Shapley, 1994), and (2) under NTU, with the additional assumption that links are strategic complements, the model is a supermodular game (Milgrom and Roberts, 1990), so the existence of equilibrium follows from the fixed-point theorem for isotone mappings (Topkis, 1979). Detailed proofs are given in the appendix.

Proposition 2.1 *Suppose that the utility function is as in (2). Under TU, for any function* u *and any constants* γ_1 *and* γ_2 *in* $\mathbb{R}$*, there is no closed cycle, so a PS network must exist.*

Proposition 2.2 *Suppose that the utility function is as in (2). Under NTU, for any function* u *and any constants* $\gamma_1 \geq 0$ *and* $\gamma_2 \geq 0$*, there is no closed cycle, so a PS network must exist.*

Remark 2.1 *The existence results in Propositions 2.1-2.2 can be extended to generalizations of the utility specification in (2) where* γ_1 *and* γ_2 *depend on the attributes. Suppose that the coefficients of* $G_{ij}G_{jk}$ *and* $G_{ij}G_{ik}G_{jk}$ *in (2) take the form of* $\gamma_1(X_i, X_j, X_k)$ *and* $\gamma_2(X_i, X_j, X_k)$*, respectively. If* $\gamma_1(X_i, X_j, X_k)$ *is symmetric in* X_i *and* X_k*, and* $\gamma_2(X_i, X_j, X_k)$ *is symmetric in* X_i*,* X_j*, and* X_k*, one can show that the result in Proposition 2.1 remains satisfied. Furthermore, the result in Proposition 2.2 holds if* $\gamma_1(X_i, X_j, X_k) \geq 0$ *and* $\gamma_2(X_i, X_j, X_k) \geq 0$ *for all* X_i*,* X_j*, and* X_k*.*

Remark 2.2 *There are other equilibrium concepts in the network literature, and they differ mainly in the coordination that individuals are assumed to have. The simplest concept is Nash equilibrium, which allows for no coordination. In the mutual-consent setting, Nash equilibrium is not appropriate because even if forming a link is beneficial for both individuals involved, it can still be optimal in the Nash sense that they do not form the link, merely due to coordination failure.*[9] *This is why Jackson and Wolinsky proposed pairwise stability, which allows two individuals to coordinate so they do not fail to form a link if that is beneficial for both. Pairwise stability only allows for the coordination of a pair on one link. There are other equilibrium concepts*

[9] This is because if i rejects the link, it does not matter whether or not j rejects it. Then rejection is a (weakly) optimal choice for j. Moreover, given j's rejection, it is also (weakly) optimal for i to reject the link.

that allow for higher-level coordination. For example, bilateral equilibrium allows for the coordination of a pair on more than one link (Goyal and Vega-Redondo (2007)), and strong stability allows for the coordination of a coalition (Dutta and Mutuswami (1997), Jackson and van den Nouweland (2005)). These concepts refine pairwise stability with further restrictions. In this paper, we want to keep the assumptions as weak as possible, so we only assume pairwise stability.

3 Partial Identification

In this section, we examine the general framework that we use to identify the model. After introducing the data generating process, we discuss multiple equilibria, the main problem in identification. Then we show how much we can learn about the parameters without imposing any restrictions on the equilibrium selection.

We consider the following data generating process. Let n be an integer generated from a distribution on $\{2, 3, \ldots\}$. We draw n individuals at random from a super-population. Each individual i is associated with a vector of attributes $X_{n,i}$ and a vector of preferences $\varepsilon_{n,i}$. We let these n individuals form links, and a PS network $G_n = (G_{n,ij})_{i \neq j}$ emerges. For notational convenience, we define $X_{n,ij} = (X_{n,i}, X_{n,j})$ to be the attributes of a pair (i, j) and $X_n = (X_{n,ij})_{i \neq j}$ the attribute profile of all the pairs. We observe the network G_n, the attribute profile X_n, but not the preference profile $\varepsilon_n = (\varepsilon_{n,ij})_{i \neq j}$. This network generating procedure is repeated independently T times, and we obtain an i.i.d. sample of networks and attribute profiles (G_{n_t}, X_{n_t}), $t = 1, \ldots, T$.

Throughout the paper we make the following assumptions.

Assumption 1 (Data generating process) *(i) We have an i.i.d. sample of* (G_{n_t}, X_{n_t})*,* $t = 1, \ldots, T$*. Let* $T \to \infty$*. (ii)* X_{n_t} *and* ε_{n_t} *are independent for all* $t = 1, \ldots, T$*. (iii)* $\varepsilon_{n_t,ij}$ *for all* $i \neq j$ *and* $t = 1, \ldots, T$ *are i.i.d. from a distribution with CDF* $F(\varepsilon_{ij}; \theta_\varepsilon)$ *supported on* $\mathbb{R}$ *that is absolutely continuous with respect to the Lebesgue measure.* $F(\varepsilon_{ij}; \theta_\varepsilon)$ *is continuously differentiable in the finite-dimensional parameter* $\theta_\varepsilon \in \Theta_\varepsilon$*.*

Assumption 2 (Utility) *The marginal utility of* i *from forming a link with* j *has a form* $\Delta U_{ij}(G_{n,-ij}, X_{n,ij}, \varepsilon_{n,ij}; \theta_u)$ *as specified in (3), where* $\theta_u = (\beta, \gamma) \in \Theta_u$ *denotes the utility parameter.*

The parameter of interest is $\theta = (\theta_u, \theta_\varepsilon) \in \Theta_u \times \Theta_\varepsilon = \Theta$.

For a given attribute profile X_n and preference profile ε_n, the model yields a collection of PS networks, denoted by $\mathcal{PS}(\Delta U_n(X_n, \varepsilon_n))$, where $\Delta U_n(X_n, \varepsilon_n) = \{\{\Delta U_{ij}(G_{n,-ij}, X_{n,ij}, \varepsilon_{n,ij})\}_{G_{n,-ij} \in \mathcal{G}_{n,-ij}}\}_{i \neq j} \in \mathbb{R}^{n(n-1)|\mathcal{G}_n|/2}$ is the marginal-utility profile, and $\mathcal{G}_{n,-ij}$ and $\mathcal{G}_n$ are the sets of all possible $G_{n,-ij}$ and G_n respectively. To complete the model, suppose there is an equilibrium selection mechanism that selects a network from the collection of PS networks. Let $\lambda_n(g_n|\mathcal{PS}(\Delta U_n(X_n, \varepsilon_n)))$ be the probability with which a network g_n is selected from the PS collection $\mathcal{PS}(\Delta U_n(X_n, \varepsilon_n))$. Then conditional on X_n the probability that we observe the network g_n is

$$\Pr(G_n = g_n | X_n) = \int \lambda_n(g_n|\mathcal{PS}(\Delta U_n(X_n, \varepsilon_n)))\, dF(\varepsilon_n) \tag{6}$$

Equation (6) is similar to what Ciliberto and Tamer (2009) establish in entry games and Bajari, Hong, and Ryan (2010) in discrete games with complete information.

Since the equilibrium selection probability in (6) is unknown when there are multiple equilibria, whether the true parameter value θ_0 can be point identified from the restriction in (6) depends on whether there is an unique equilibrium. If for any $\theta \in \Theta$ there is a network that can only be a unique equilibrium, then under certain conditions the unique equilibrium may provide moment restrictions to point identify θ_0. However, if for some $\theta \in \Theta$ all the networks are part of multiple equilibria, then θ_0 cannot be point identified without additional restrictions on the equilibrium selection. In this case, we encounter the incomplete problem addressed in the literature (Bresnahan and Reiss (1991), Tamer (2003)).

For the network formation game described in Section 2, the presence of multiple equilibria is prevalent because of the interdependence of marginal utilities across links.[10] We illustrate multiple equilibria in Example 3.1.

Example 3.1 *Consider networks of size $n = 3$. Figure 1 shows the eight possible network configurations. Consider the utility function as in (2) with $u(X_i, X_j; \beta) = u(X_j, X_i; \beta)$, $\gamma_1 > 0$, and $\gamma_2 > 0$. Abbreviate $u(X_i, X_j; \beta)$ as u_{ij}. For simplicity we assume $\varepsilon_{ij} = \varepsilon_{ji}$, so $\varepsilon = (\varepsilon_{12}, \varepsilon_{23}, \varepsilon_{13}) \in \mathbb{R}^3$. Given the utility specification, we calculate all possible collections of PS networks under TU. The regions of ε that correspond to each collection of PS networks are presented in Figure 2, where a network*

[10] Note that if there is no utility interdependence, i.e., $\Delta U_{ij}(G_{n,-ij}, X_{n,ij}, \varepsilon_{n,ij}) = \Delta U_{ij}(X_{n,ij}, \varepsilon_{n,ij})$, then a pairwise stable network must be unique.

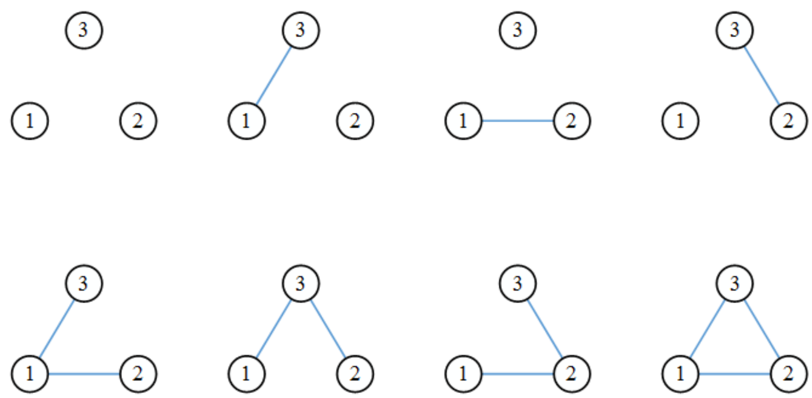

Figure 1: Networks of Three Individuals

g is represented by the vector $(g_{12}, g_{23}, g_{13}) \in \{0,1\}^3$. In this example, all the eight networks belong to certain multiple equilibria; no network can be a unique equilibrium.

One can achieve point identification by making certain assumptions about the equilibrium selection. See Remark 3.1 for a detailed discussion. In this paper, we do not want to impose any restrictions on the equilibrium selection, so we get around the non-identifiability issue using partial identification. This approach has been widely applied to game-theoretic models with multiple equilibria (Andrews, Berry and Jia (2004), Pakes, Porter, Ho and Ishii (2006), Berry and Tamer (2006), Ciliberto and Tamer (2009)).

Following closely Ciliberto and Tamer (2009), we divide the integral in (6) into two parts, depending on whether there is a unique equilibrium or multiple equilibria,

$$\begin{aligned}\Pr\left(G_n = g_n \middle| X_n\right) = & \int_{g_n \in \mathcal{PS}(\Delta U_n(X_n,\varepsilon_n)) \& |\mathcal{PS}(\Delta U_n(X_n,\varepsilon_n))|=1} dF\left(\varepsilon_n\right) \\ & + \int_{g_n \in \mathcal{PS}(\Delta U_n(X_n,\varepsilon_n)) \& |\mathcal{PS}(\Delta U_n(X_n,\varepsilon_n))|\geq 2} \lambda_n\left(g_n | \mathcal{PS}\left(\Delta U_n\left(X_n, \varepsilon_n\right)\right)\right) dF\left(\varepsilon_n\right),\end{aligned} \tag{7}$$

Note that the selection probability is trivially 1 when a network is a unique equilibrium. When there are multiple equilibria, the selection probability, though unknown, lies between 0 and 1. Replacing the selection probability with these bounds, we derive an upper and lower bound for $\Pr\left(G_n = g_n \middle| X_n\right)$, i.e.,

$$\Pr\left(G_n = g_n \middle| X_n\right) \leq \int_{g_n \in \mathcal{PS}(\Delta U_n(X_n,\varepsilon_n))} dF\left(\varepsilon_n\right), \tag{8}$$

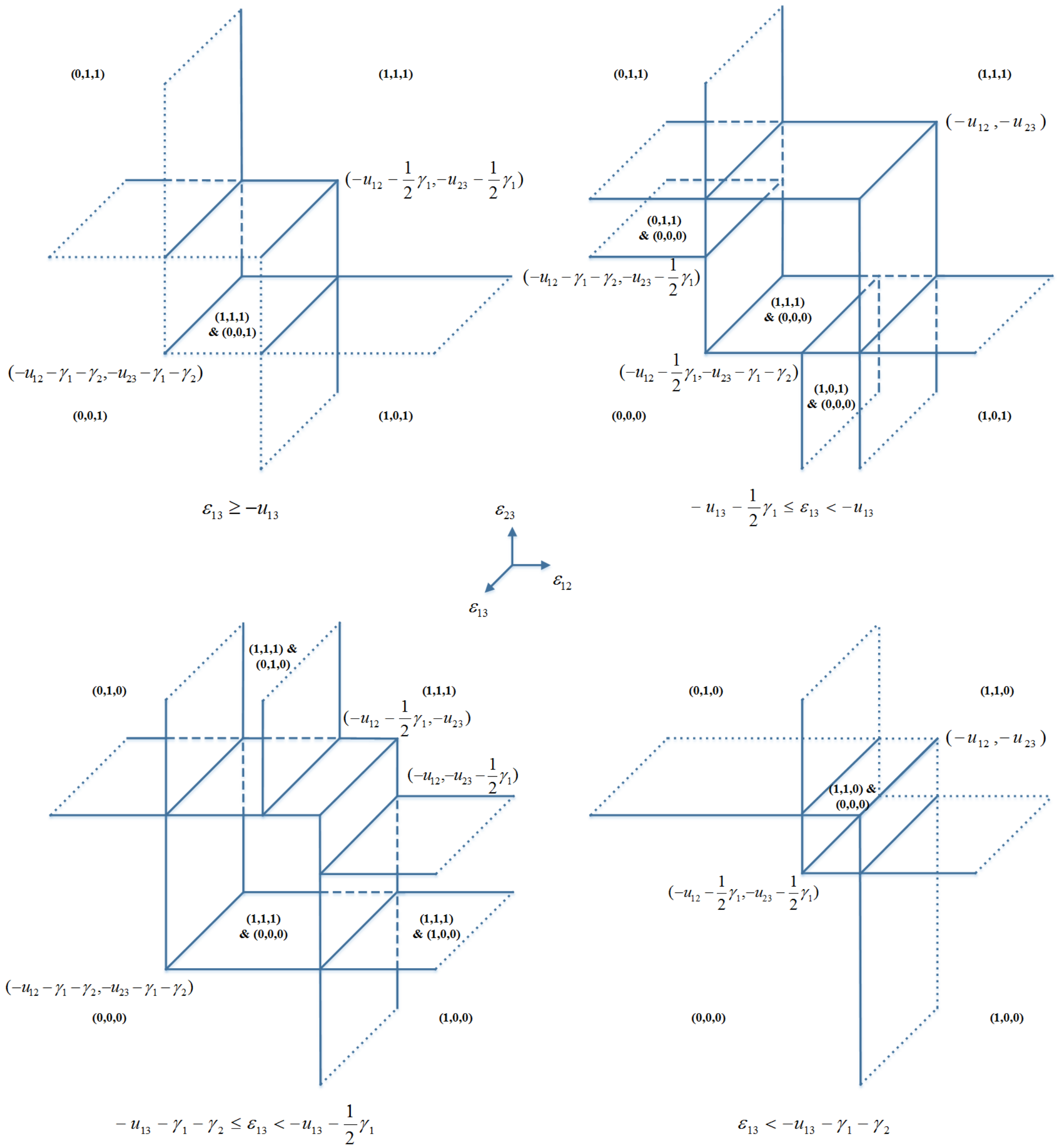

Figure 2: All Possible Equilibria and the Partition of the ε Space

and

$$\Pr\left(G_n = g_n \middle| X_n\right) \geq \int_{g_n \in \mathcal{PS}(\Delta U_n(X_n,\varepsilon_n)) \& |\mathcal{PS}(\Delta U_n(X_n,\varepsilon_n))|=1} dF\left(\varepsilon_n\right). \tag{9}$$

The upper bound is the probability that network g_n is PS, and the lower bound is the probability that network g_n is uniquely PS. These are the best possible bounds for $\Pr\left(G_n = g_n \middle| X_n\right)$ because the selection probability in (7) can be any value between 0 and 1.

Unfortunately, these bounds suffer from the curse of dimensionality in large networks. In particular, the lower bound in (9) is computationally infeasible if n is large. This is because to compute the lower bound, we need to check pairwise stability for $2^{n(n-1)/2}$ possible networks.[11] This is computationally intractable even for a moderate value n. For example, in the case of 20 people, the number of possible networks is $2^{190} \approx 10^{57}$.

Remark 3.1 *An alternative approach is to achieve point identification by making additional assumptions about the equilibrium selection. In network formation, one way to do this is to consider a sequential model as in Jackson and Watts (2002) (see also Christakis et al. (2010) and Mele (2011)). This sequential model assumes that individuals are myopic and form links in a random sequence: in each period only one pair of individuals is randomly selected and only that pair can update their relationship. The sequence of networks realized in each period form a Markov chain with states corresponding to the networks. Under certain conditions*[12] *the Markov chain converges to a unique stationary distribution, which typically assigns probability one to a single PS network.*[13] *Hence the stationary distribution amounts to a particular selection rule. Alternatively, one can assume a more general equilibrium selection mechanism, for example, by specifying a parametric form (Bajari, Hong, and Ryan (2010)) or considering a nonparametric equilibrium selection (Bajari, Hahn, Hong, and Ridder (2011)). Note that in the game we consider a fully nonparametric equilibrium selection is not identified. Certain restrictions must be imposed on it in order*

[11] Unlike the upper bound, the lower bound has no closed form and needs to be computed by simulation. For each simulated ε_n, we need to check whether a network is uniquely pairwise stable, which amounts to checking pairwise stability for all possible networks.

[12] An example of such conditions would be (i) the individuals are assumed to make mistakes (i.e., forming or deleting a link randomly rather than based on utility maximization) and (ii) the probability of making a mistake is sufficiently small.

[13] This network is essentially the most "stable" one among all the PS networks, or more precisely, the network that has the minimum resistance (Young (1993), Jackson and Watts (2002)).

to achieve identification.

4 Partial Identification from Subnetworks

4.1 Inequalities from Subnetworks

We propose a novel approach to reduce the dimensionality of the problem. The idea is to derive bounds for certain parts of a network, called subnetworks. A subnetwork is the restriction of a network to a subset of the individuals. To be precise, let G_n be a network of n nodes. For any subset $A \subseteq [n]$, we say $G_{n,A}$ is the *subnetwork* of G_n in A if it consists of the edges in G_n that connect two nodes in A, i.e., $G_{n,A} = (G_{n,ij})_{i,j\in A, i\neq j}$. Moreover, we define $G_{n,-A}$ to be the *complement* of $G_{n,A}$, i.e., the remainder of G_n after the edges in $G_{n,A}$ are deleted. It consists of the edges in G_n that connect either two nodes in $A^c = [n]\backslash A$ or one node in A and another in A^c, i.e., $G_{n,-A} = (G_{n,ij})_{i\notin A \cup j\notin A, i\neq j}$. In matrix notation, the subnetwork $G_{n,A}$ corresponds to the submatrix of G_n with rows and columns in A, and its complement $G_{n,-A}$ is the remainder of G_n after the submatrix in A is deleted. The sets of all possible $G_{n,A}$ and $G_{n,-A}$ are denoted by $\mathcal{G}_{n,A}$ and $\mathcal{G}_{n,-A}$.

For any fixed subset $A \subseteq [n]$, it is clear from the decomposition $G_n = (G_{n,A}, G_{n,-A})$ that the distribution of the subnetwork $G_{n,A}$ is simply a marginal distribution of the network G_n. That is, conditional on X_n the probability of observing a subnetwork $g_{n,A}$ is

$$\begin{aligned}\Pr\left(G_{n,A} = g_{n,A} \middle| X_n\right) &= \sum_{g_{n,-A}} \Pr\left(G_{n,A} = g_{n,A}, G_{n,-A} = g_{n,-A} \middle| X_n\right) \\ &= \int \sum_{g_{n,-A}} \lambda_n\left(g_{n,A}, g_{n,-A} \middle| \mathcal{PS}\left(\Delta U_n\left(X_n, \varepsilon_n\right)\right)\right) dF\left(\varepsilon_n\right). \quad (10)\end{aligned}$$

The summed equilibrium selection probability in (10) is unknown unless all the networks in $\mathcal{PS}\left(\left(\Delta U_n\left(X_n, \varepsilon_n\right)\right)\right)$ have the same subnetwork in A. Following the same idea as in the previous section, we can derive an upper and lower bound for $\Pr\left(G_{n,A} = g_{n,A} \middle| X_n\right)$. Specifically, divide the integral in (10) into two parts, depending on whether the PS networks have a unique subnetwork or multiple subnetworks,

$$
\begin{aligned}
\Pr\left(G_{n,A}=g_{n,A}\right|X_n) = & \int_{\substack{g_{n,A}\in\mathcal{PS}_A(\Delta U_n(X_n,\varepsilon_n))\\ \&|\mathcal{PS}_A(\Delta U_n(X_n,\varepsilon_n))|=1}} dF\left(\varepsilon_n\right) \\
& + \int_{\substack{g_{n,A}\in\mathcal{PS}_A(\Delta U_n(X_n,\varepsilon_n))\\ \&|\mathcal{PS}_A(\Delta U_n(X_n,\varepsilon_n))|\geq 2}} \sum_{g_{n,-A}} \lambda_n\left(g_{n,A},g_{n,-A}|\mathcal{PS}\left(\Delta U_n\left(X_n,\varepsilon_n\right)\right)\right) dF\left(\varepsilon_n\right),
\end{aligned}
\tag{11}
$$

where $\mathcal{PS}_A\left(\Delta U_n\left(X_n,\varepsilon_n\right)\right)=\{g_{n,A}\in\mathcal{G}_{n,A}:\exists g_{n,-A}\in\mathcal{G}_{n,-A},(g_{n,A},g_{n,-A})\in\mathcal{PS}(\Delta U_n(X_n,\varepsilon_n))\}$ is the set of subnetworks in A that are part of a network in $\mathcal{PS}\left(\Delta U_n\left(X_n,\varepsilon_n\right)\right)$. Replacing the sum term in (11) by 0 and 1 yields

$$
H_{2n}\left(g_{n,A},X_n\right)\leq \Pr\left(G_{n,A}=g_{n,A}\right|X_n)\leq H_{1n}\left(g_{n,A},X_n\right) \tag{12}
$$

where

$$
\begin{aligned}
H_{1n}\left(g_{n,A},X_n\right) &= \int_{g_{n,A}\in\mathcal{PS}_A(\Delta U_n(X_n,\varepsilon_n))} dF\left(\varepsilon_n\right)\\
H_{2n}\left(g_{n,A},X_n\right) &= \int_{g_{n,A}\in\mathcal{PS}_A(\Delta U_n(X_n,\varepsilon_n))\&|\mathcal{PS}_A(\Delta U_n(X_n,\varepsilon_n))|=1} dF\left(\varepsilon_n\right)
\end{aligned}
$$

These bounds are analogous to those network bounds in (8) and (9): the upper bound in (12) is the probability that $g_{n,A}$ is part of a PS network, and the lower bound in (12) is the probability that only $g_{n,A}$ is part of a PS network. These are the best possible bounds for $\Pr\left(G_{n,A}=g_{n,A}\right|X_n)$ because the summed selection probability in (11) can be any value between 0 and 1. In contrast to the lower bound in (9), these bounds can be computed even in large networks as long as the subnetworks are chosen to be small. Details about the computation are discussed in Section 6.

Example 4.1 (Example 3.1 continued) *Assume the same setting as in Example 3.1. We calculate the upper and lower bounds in (12) for subnetwork $G_{12}=1$ (suppress the subscript n). Note that the complement of the subnetwork G_{-12} takes four possible values $\{(1,1),(1,0),(0,1),(0,0)\}$. The regions in Figure 2 in which $G_{12}=1$ associated with any of these complements is PS gives the upper bound, i.e.,*

$$
\Pr\left(G_{12}=1\right|X) \leq \int_{\{(1,1,1)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF\left(\varepsilon\right)+\int_{\{(1,1,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF\left(\varepsilon\right)
$$

$$
\begin{aligned}
&+\int_{\{(1,0,1)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon)+\int_{\{(1,0,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon) \\
&+\int_{\{(1,1,1),(1,0,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon)+\int_{\{(1,1,1),(0,1,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon) \\
&+\int_{\{(1,1,1),(0,0,1)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon)+\int_{\{(1,1,1),(0,0,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon) \\
&+\int_{\{(1,1,0),(0,0,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon)+\int_{\{(1,0,1),(0,0,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon).
\end{aligned}
$$

The lower bound can be derived from the subset of the regions in the upper bound where $G_{12}=0$ associated with any complement is not PS, i.e.,

$$
\begin{aligned}
\Pr\left(G_{12}=1|\,X\right) \geq\ & \int_{\{(1,1,1)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon)+\int_{\{(1,1,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon) \\
&+\int_{\{(1,0,1)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon)+\int_{\{(1,0,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon) \\
&+\int_{\{(1,1,1),(1,0,0)\}=\mathcal{PS}(\Delta U(X,\varepsilon))} dF(\varepsilon).
\end{aligned}
$$

4.2 Convergence of the Subnetwork Inequalities

A major concern about the bounds from subnetworks is their performance when networks are large. In order for these bounds to be useful in large networks, they must remain informative and provide nontrivial restrictions for the parameters as n tends to infinity. We also want the inequality restrictions in (12) to be convergent as n increases so the inference of the parameters is robust to the size of networks. These features require that both the subnetwork choice probabilities and their bounds converge to some nontrivial limits as n approaches infinity. Our objective in this section is to show that under mild assumptions on the equilibrium selection these asymptotic properties are actually satisfied.

The convergence of the subnetwork choice probabilities is achieved under assumptions on the equilibrium selection that are motivated by the convergence of exchangeable random graphs (Lovász and Szegedy, 2006; Diaconis and Janson, 2008; Lovász, 2012). Exchangeability is relevant in our context because the utility specification in (2) does not depend on how we label the individuals, so if the equilibrium selection mechanism is also assumed not to depend on the labels, the distribution of networks is invariant under permutations of the labels and is thus exchangeable. Note that because individuals have attributes, only the individuals with the same attributes are exchangeable. We define such restricted exchangeability by considering a network G_n together with its attribute profile X_n and calling (G_n, X_n) an attributed network or simply network. Recall that (G_n, X_n) is a $n \times n$ matrix on $\{0,1\} \times \mathcal{X}^2$. Denote the set of all possible (G_n, X_n) by $(\mathcal{G}_n, \mathcal{X}_n)$. For any network $(G_n, X_n) \in (\mathcal{G}_n, \mathcal{X}_n)$, we define $(G_n^\pi, X_n^\pi) \in (\mathcal{G}_n, \mathcal{X}_n)$ to be the network induced by π, where π is a permutation over $[n]$. This is the network obtained by permuting the rows and columns of (G_n, X_n) according to π, so the (i,j) element of (G_n^π, X_n^π) is equal to the $(\pi(i), \pi(j))$ element of (G_n, X_n) for all $i \neq j$. For a network with infinite number of individuals $(G, X) = (G_{ij}, X_{ij})_{i,j \geq 1, i \neq j} \in (\mathcal{G}_\infty, \mathcal{X}_\infty)$, we also let $(G^\pi, X^\pi) \in (\mathcal{G}_\infty, \mathcal{X}_\infty)$ be the infinite network induced by π, where π is a permutation over $\mathbb{N} = \{1, 2, \ldots\}$ that leaves all but a finite number of terms fixed. Exchangeability means that all the networks induced by permutations have the same distribution as the original network.

Definition 4.1 *(i) A finite network (G_n, X_n) is exchangeable if for any permutation π over $[n]$ the induced network (G_n^π, X_n^π) has the same distribution as (G_n, X_n), i.e., $(G_n^\pi, X_n^\pi) \overset{d}{=} (G_n, X_n)$. (ii) An infinite network (G, X) is exchangeable if for any permutation π over $\mathbb{N}$ that permutes a finite number of elements in $\mathbb{N}$, we have $(G^\pi, X^\pi) \overset{d}{=} (G, X)$.*

It is easy to show that if the equilibrium selection mechanism λ_n is invariant under permutations of labels, i.e., for any permutation π over $[n]$, any network (g_n, x_n), and any preference profile ε_n, we have

$$\lambda_n \left(g_n^\pi \middle| \mathcal{PS} \left(\Delta U_n \left(x_n^\pi, \varepsilon_n^\pi \right) \right) \right) = \lambda_n \left(g_n \middle| \mathcal{PS} \left(\Delta U_n \left(x_n, \varepsilon_n \right) \right) \right), \tag{13}$$

where ε_n^π is the preference profile induced by the permutation π, then the network (G_n, X_n) generated by the equilibrium selection λ_n is exchangeable. We impose this restriction in Assumption 3(i).

The exchangeability of (G_n, X_n) has two immediate implications. First, for any subsets A and $A' \subseteq [n]$ with the same size $|A| = |A'|$, the subnetworks $(G_{n,A}, X_{n,A})$ and $(G_{n,A'}, X_{n,A'})$ have the same distribution and thus the same choice probabilities, i.e., $\Pr\left(G_{n,A} = g_a \middle| X_{n,A} = x_a, X_{n,-A}\right) = \Pr\left(G_{n,A'} = g_a \middle| X_{n,A'} = x_a, X_{n,-A'}\right)$. Hence it suffices to consider the subnetwork in $A = [a]$, denoted by $(G_{n,a}, X_{n,a})$, and its choice probabilities $\Pr\left(G_{n,a} = g_a \middle| X_{n,a} = x_a, X_{n,-a}\right)$. Second, if a subnetwork in $[a]$

has two individuals with the same X, then due to indeterminacy in the labeling the links and attributes in $[a]$ may be represented as different subnetworks that are isomorphic. We say that two subnetworks (g_a, x_a) and (g'_a, x'_a) are *isomorphic* if there is a permutation π over $[a]$ such that $g'_{a,ij} = g_{a,\pi(i)\pi(j)}$ and $x'_{a,ij} = x_{a,\pi(i)\pi(j)}$ for $i, j \leq a$, $i \neq j$, i.e., (g'_a, x'_a) is induced from (g_a, x_a) by π. Exchangeability implies that the choice probabilities evaluated at such isomorphic subnetworks are the same, i.e., $\Pr\left(G_{n,a} = g_a \middle| X_{n,a} = x_a, X_{n,-a}\right) = \Pr\left(G_{n,a} = g'_a \middle| X_{n,a} = x'_a, X_{n,-a}\right)$. Thus we can resolve the indeterminacy by defining the "=" symbol in the choice probability as an isomorphism and viewing $\Pr\left(G_{n,a} = g_a \middle| X_{n,a} = x_a, X_{n,-a}\right)$ as a function of unlabeled subnetworks (g_a, x_a), i.e., the equivalence classes of subnetworks defined by the isomorphism relation. Note that the use of isomorphism and unlabeled subnetworks is needed only for discrete X. If X is continuous, two individuals have distinct X with probability 1, so no subnetworks are isomorphic and labeled and unlabeled subnetworks are the same.

In the graph limit theory, the convergence of exchangeable random graphs is defined in terms of the convergence of subgraph densities (Lovász and Szegedy, 2006; Diaconis and Janson, 2008). Motivated by this insight, we further restrict the sequence of equilibrium selection mechanisms $\{\lambda_n, n \geq 2\}$ so that the finite exchangeable networks generated by $\{\lambda_n, n \geq 2\}$ converge to an infinite exchangeable network as $n \to \infty$, thereby yielding the desired convergence of the subnetwork probabilities. To be precise, for any fixed subnetwork $(g_a, x_a) \in (\mathcal{G}_a, \mathcal{X}_a)$ and any given (finite or infinite) network (G, X), we define the *subnetwork density* $t_{ind}\left((g_a, x_a), (G, X)\right)$ as the probability that a randomly selected subset A in the node set of G with size $|A| = a$ yields a subnetwork (G_A, X_A) that equals (g_a, x_a) (in the sense of isomorphism), i.e., $t_{ind}\left((g_a, x_a), (G, X)\right) = \Pr\left(G_{n,A} = g_a, X_{n,A} = x_a \middle| G, X\right)$. For a finite network (G_n, X_n) with $n \geq a$, the subnetwork density is given by $t_{ind}\left((g_a, x_a), (G_n, X_n)\right) = \frac{1}{\binom{n}{a}} \sum_{A \subseteq [n]:|A|=a} 1\left\{G_{n,A} = g_a, X_{n,A} = x_a\right\}$. We say a sequence of finite networks converges to an infinite network if all the subnetwork densities of the finite networks converge to those of the infinite network. Note that for finite exchangeable networks the limit network must also be exchangeable.[14]

Definition 4.2 *A sequence of finite exchangeable networks* $\{(G_n, X_n), n \geq 2\}$ *con-*

[14] This is because the subnetwork density of a finite exchangeable network $t_{ind}\left((g_a, x_a), (G_n, X_n)\right)$ depends only on the isomorphism type of the subnetwork (g_a, x_a), so does the subnetwork density of the limit network $t_{ind}\left((g_a, x_a), (G^*, X^*)\right)$.

verge to an infinite exchangeable network $(G^*, X^*) = \left(G^*_{ij}, X^*_{ij}\right)_{i,j\geq 1, i\neq j}$ *if for any* $a \leq n$ *and any subnetwork* $(g_a, x_a) \in (\mathcal{G}_a, \mathcal{X}_a)$*, the random variable* $t_{ind}\left((g_a, x_a), (G_n, X_n)\right)$ *converges in distribution to the random variable* $t_{ind}\left((g_a, x_a), (G^*, X^*)\right)$ *as* $n \to \infty$*.*

Under Assumption 1 a sequence of attribute profiles $\{X_n,\ n \geq 2\}$ can be embedded into an infinite exchangeable array $X^* = \left(X^*_{ij}\right)_{i,j\geq 1, i\neq j}$ with $X^*_{ij} = X_{n,ij}$ for all $n \geq 2$ and all $i, j \leq n$, $i \neq j$. The convergence in Definition 4.2 then amounts to a restriction on the equilibrium selection so that $\{G_n,\ n \geq 2\}$ can be "asymptotically embedded" into G^*. We impose this restriction in Assumption 3(ii). This assumption rules out equilibrium selection mechanisms that may oscillate between pairwise stable networks with different subnetwork densities as $n \to \infty$. For example, for the utility specification in (2) with γ_1, $\gamma_2 > 0$, if for any (X_n, ε_n) the equilibrium selection mechanism λ_n selects from $\mathcal{PS}\left(\Delta U\left(X_n, \varepsilon_n\right)\right)$ the largest network when n is odd and the smallest network when n is even, then the sequence of networks generated by such λ_n can never converge.[15]

Assumption 3 *The sequence of equilibrium selection mechanisms* $\lambda_n : \mathcal{G}_n \times 2^{\mathcal{G}_n} \to [0,1]$*,* $n \geq 2$*, satisfies that (i) for any* $n \geq 2$*,* λ_n *is invariant under permutations of labels, i.e., the condition in (13) holds, and (ii) the sequence of networks* $\{(G_n, X_n), n \geq 2\}$ *generated by* $\{\lambda_n, n \geq 2\}$ *converges to an infinite exchangeable network* $(G^*, X^*) = \left(G^*_{ij}, X^*_{ij}\right)_{i,j\geq 1, i\neq j}$ *as* $n \to \infty$*.*

The network in the limit (G^*, X^*) is an exchangeable infinite two-dimensional array. From the Aldous-Hoover theorem (e.g., Kallenberg (2005), Theorem 7.22), it has a representation

$$\left(G^*_{ij}, X^*_{ij}\right) = f\left(\xi_0, \xi_i, \xi_j, \xi_{ij}\right) \text{ a.s.,} \quad \forall i, j \geq 1, i \neq j \tag{14}$$

for a measurable function $f : [0,1]^4 \to \{0,1\} \times \mathcal{X}^2$ that is symmetric in ξ_i and ξ_j and some i.i.d. random variables $(\xi_i)_{i\geq 0}$ and $\left(\xi_{ij}\right)_{i,j\geq 1, i\neq j}$ with $\xi_{ij} = \xi_{ji}$ that are uniformly distributed on $[0,1]$. We can further define a function $W : [0,1]^3 \to [0,1]$ as $W\left(\xi_0, \xi_i, \xi_j\right) = \Pr\left(f_1\left(\xi_0, \xi_i, \xi_j, \xi_{ij}\right) = 1 \middle| \xi_0, \xi_i, \xi_j\right)$, where f_1 is the component of

[15] When the model is a supermodular game, any collection of PS networks has the largest and smallest elements, i.e., there exist PS networks g^0 and g^1 such that $g^0 \leq g \leq g^1$ for any PS network g, where "$\leq$" means element-wise smaller than or equal to.

f that corresponds to G^*_{ij}, so the links in G^* can be equivalently represented as

$$G^*_{ij} = 1\left\{W\left(\xi_0, \xi_i, \xi_j\right) \geq \xi_{ij}\right\} \text{ a.s.,} \quad \forall i, j \geq 1, i \neq j \tag{15}$$

for some i.i.d. $U(0,1)$ random variables independently of $(\xi_i)_{i\geq 0}$, which we still denote by $\left(\xi_{ij}\right)_{i,j\geq 1, i\neq j}$. The function W in (15) is called a graphon (Lovász and Szegedy, 2006). While the links are dependent as a result of the pairwise stability condition and equilibrium selection mechanism, the representation in (15) indicates that the dependence has a particular structure such that conditional on some network heterogeneity ξ_0 and individual heterogeneity $(\xi_i)_{i\geq 1}$, the links become independent. This conditional independence feature is useful in analyzing the asymptotic properties of link frequencies and subnetwork probabilities. Note that the function W in (15) must satisfy $W \not\equiv 0$. Otherwise, we obtain an empty network with probability 1, which cannot be a limit of the networks generated from the data generating process we consider.

Under the assumptions on the equilibrium selection, we show in Theorem 4.1 that the subnetwork probabilities in an n-player network converge to the subnetwork probabilities in the limit network as $n \to \infty$. Moreover, this implies that the average numbers of friends and friends in common converge as $n \to \infty$, so the normalization rate in the utility specification in (3) is appropriate.

Theorem 4.1 *Let $\{(G_n, X_n), n \geq 2\}$ be a sequence of networks that satisfies Assumptions 1 and 3. For any $a \leq n$ and any $(g_a, x_a) \in (\mathcal{G}_a, \mathcal{X}_a)$,*

$$\Pr\left(G_{n,a} = g_a \middle| X_{n,a} = x_a, X_{n,-a}\right) \xrightarrow{a.s.} \Pr\left(G^*_a = g_a \middle| X^*_a = x_a, X^*_{-a}\right)$$

as $n \to \infty$.

Proof. See the appendix. ■

Corollary 4.2 *Let $\{(G_n, X_n), n \geq 2\}$ be a sequence of networks that satisfies Assumptions 1 and 3. For any $i, j \leq n$, $i \neq j$, and an arbitrary $k \neq i, j$, we have*

$$\frac{1}{n-2}\sum_{k'\neq i,j} G_{n,ik'} \xrightarrow{d} \mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_k\right)\middle| \xi_0, \xi_i\right]$$

and

$$\frac{1}{n-2}\sum_{k'\neq i,j} G_{n,ik'}G_{n,jk'} \xrightarrow{d} \mathbb{E}\left[W\left(\xi_0,\xi_i,\xi_k\right)W\left(\xi_0,\xi_j,\xi_k\right)\middle|\,\xi_0,\xi_i,\xi_j\right]$$

as $n\to\infty$*.*

Proof. See the appendix. ■

Remark 4.1 (Dense networks) *The exchangeability conditions in Assumption 3 imply that the total number of links in a network is* $\sum_{i=1}^{n}\sum_{j=i+1}^{n} G_{n,ij} = O_p\left(n^2\right)$ *(see the appendix for a proof). Such networks are dense in the stochastic sense.*[16] *We can also see from Corollary 4.2 that the degree of an individual is* $O_p\left(n\right)$*, so the probability that an individual is isolated approaches zero. It may be possible to extend our approach to sparse networks (which have* $o\left(n^2\right)$ *links), but this is beyond the scope of the paper. See Menzel (2016b) for work on strategic network formation with sparsity.*

Remark 4.2 (Continuous X) *Our definition of the subnetwork densities follows closely the subgraph densities defined in the graph limit theory for graphs without attributes (Lovász and Szegedy, 2006; Lovász, 2012). This definition assumes implicitly that* X *is discrete, which simplifies the exposition but is unnecessary and can be relaxed. In fact, if* X *is continuous, we can generalize the subnetwork density of network* (G,X) *to be* $t_{ind}\left(\left(g_a,C_a\right),\left(G,X\right)\right)=\Pr\left(G_A=g_a,X_A\in C_a\middle|\,G,X\right)$ *where* C_a *is a Borel subset of* $\mathcal{X}_a$*. Suppose that Assumption 3 is satisfied with the convergence condition defined by this generalized subnetwork density. We show in the appendix that the results in Theorem 4.1 and Corollary 4.2 still hold.*

Now we examine the bounds in (12). Under Assumption 1 these bounds are invariant under permutations of labels, so subnetworks in any two subsets $A,A'\subseteq[n]$ with $|A|=|A'|$ and $X_{n,A}=X_{n,A'}$ have the same bounds for all $g_{n,A}=g_{n,A'}$. It is thus sufficient to consider the subnetwork in $[a]$ and its bounds, which we denote by $H_{1n}\left(g_a,x_a,X_{n,-a}\right)$ and $H_{2n}\left(g_a,x_a,X_{n,-a}\right)$. In contrast to the network bounds in (8) and (9) which vanish to 0 as $n\to\infty$, Lemma 4.3 indicates that these bounds for any fixed a are bounded away from 0 and 1. More importantly, they also converge to some limits as $n\to\infty$, as proved in Theorem 4.4.

[16] A network with $\Theta\left(n^2\right)$ links is called a dense network (Bollobas and Riordan, 2009), where $Y_n=\Theta\left(n^2\right)$ if there are $c_1,c_2>0$ such that $c_1\leq\frac{Y_n}{n^2}\leq c_2$ for n sufficiently large.

Lemma 4.3 *Let $\{(G_n, X_n), n \geq 2\}$ be a sequence of networks that satisfies Assumptions 1-2. For any $a \leq n$ and any $(g_a, x_a) \in (\mathcal{G}_a, \mathcal{X}_a)$, the bounds $H_{1n}(g_a, x_a, X_{n,-a})$ and $H_{2n}(g_a, x_a, X_{n,-a})$ in (12) satisfy*

$$\bar{H}_2(g_a, x_a) \leq H_{2n}(g_a, x_a, X_{n,-a}) \leq H_{1n}(g_a, x_a, X_{n,-a}) \leq \bar{H}_1(g_a, x_a)$$

for some deterministic functions $\bar{H}_1(g_a, x_a)$ and $\bar{H}_2(g_a, x_a)$ such that $0 < \bar{H}_2(g_a, x_a) < \bar{H}_1(g_a, x_a) < 1$.

Proof. See the appendix. ∎

Theorem 4.4 *Let $\{(G_n, X_n), n \geq 2\}$ be a sequence of networks that satisfies Assumptions 1-2. Let $X^* = (X^*_{ij})_{i,j \geq 1, i \neq j}$ be the infinite array with $X^*_{ij} = X_{n,ij}$ for all $n \geq 2$ and all $i, j \leq n$, $i \neq j$. Then for any $a \leq n$ and any $(g_a, x_a) \in (\mathcal{G}_a, \mathcal{X}_a)$, the bounds $H_{1n}(g_a, x_a, X_{n,-a})$ and $H_{2n}(g_a, x_a, X_{n,-a})$ in (12) satisfy*

$$\begin{aligned} H_{1n}(g_a, x_a, X_{n,-a}) &\overset{a.s.}{\to} H^*_1\left(g_a, x_a, X^*_{-a}\right) \\ H_{2n}(g_a, x_a, X_{n,-a}) &\overset{a.s.}{\to} H^*_2\left(g_a, x_a, X^*_{-a}\right) \end{aligned}$$

as $n \to \infty$, for some functions $H^*_1\left(g_a, x_a, X^*_{-a}\right)$ and $H^*_2\left(g_a, x_a, X^*_{-a}\right)$.

Proof. See the appendix. ∎

The upper bound $\bar{H}_1(g_a, x_a)$ in Lemma 4.3 is the probability that the subnetwork (g_a, x_a) is PS for the "most favorable" complement, and the lower bound $\bar{H}_2(g_a, x_a)$ is the probability that (g_a, x_a) is uniquely PS for the "least favorable" complement. Because the effects of friends of friends and friends in common in (3) are normalized by $n - 2$, the overall utility externality from any complement is bounded. Hence even for the extreme complements the subnetwork probabilities are bounded away from 0 and 1. Theorem 4.4 strengthens the results in the lemma by showing that the subnetwork bounds actually converge: the upper bound converges to the probability that the subnetwork (g_a, x_a) is PS for an infinite PS complement generated under the "most favorable" equilibrium selection mechanism, and the lower bound converges to the probability that (g_a, x_a) is uniquely PS for an infinite PS complement generated under the "least favorable" equilibrium selection mechanism.[17] The exact forms of the

[17] The limits of the bounds may be random due to the randomness in X^*_{-a}. In a special case when the attributes in a network are i.i.d., the limits do not depend on X^*_{-a} and reduce to deterministic functions. See the proof of the theorem for details.

bounds and limits are given in the proofs. These results together with Theorem 4.1 ensure that the subnetwork inequalities scale well as n increases, so small subnetworks can provide useful information about the parameter even in large networks.

In addition, the convergence of the bounds provides an attractive possibility to reduce the computational complexity in large networks. We can approximate the bounds in an n-player network by the bounds from an m-player network with $m \ll n$. The approximation is arbitrarily well for sufficiently large m.

Remark 4.3 *The subnetwork inequalities in (12) and their properties established in Theorem 4.1 and Lemma 4.3 do not require the utility specification in (2). They can apply to more general specifications, for example, where γ_1 and γ_2 depend on the attributes X, so long as the existence of a PS network is guaranteed as discussed in Remark 4.1. We will see in Section 6 that the computation of the bounds does not need γ_1 and γ_2 to be constant either; the computational cost remains the same when γ_1 and γ_2 depend on X.* [18]

4.3 Identified Sets

The inequalities in (12) are satisfied by the true parameter θ_0, i.e.,

$$H_{2n}(g_a, x_a, X_{n,-a}; \theta_0) \leq \Pr\left(G_{n,a} = g_a \middle| X_{n,a} = x_a, X_{n,-a}\right) \leq H_{1n}(g_a, x_a, X_{n,-a}; \theta_0) \tag{16}$$

for all (g_a, x_a), all $X_{n,-a}$ and all $a \leq n$. We define the identified set from subnetworks of size a as the collection of θ that satisfy the inequalities in (16) for that a,

$$\Theta_I(a) = \{\theta \in \Theta : (16) \text{ holds for the given } a \text{ with } \theta \text{ in place of } \theta_0\}. \tag{17}$$

and define the identified set to be $\Theta_I = \bigcap_{a=2}^{\bar{a}} \Theta_I(a)$ for some positive integer $\bar{a}$.

Remark 4.4 (Sharp Identified Sets) *The identified sets defined in (17) are not sharp. One can construct the sharp identified set for each a, denoted by $\Theta_I^s(a)$, as the collection of θ such that (10) holds for some equilibrium selection mechanism, similarly as in Beresteanu, Molchanov and Molinari (2011). The inequalities that*

[18] The convergence of the bounds (Theorem 4.4) is proved for the specification in (2), but this result is for theoretical purpose and is not needed in the estimation. It may be possible to generalize the proof to allow for attribute-dependent γ_1 and γ_2. We leave it for future research.

define $\Theta_I^s(a)$ *are of the form* $\Pr(G_{n,a} \in \mathcal{H}_a | X_n) \leq \int 1\{\mathcal{H}_a \subseteq PS_{[a]}((\Delta U_n(X_n, \varepsilon_n)))\}$ $dF(\varepsilon_n)$*, where* $\mathcal{H}_a$ *is any subset of* $\mathcal{G}_a$*. These sharp identified sets satisfy* $\Theta_I^s(a_2) \subseteq \Theta_I^s(a_1)$ *for* $a_2 > a_1$*,*[19] *so bounds from larger subnetworks provide more information about the parameter.*[20] *One can show that the convergence results in this section also hold for the inequalities defining the sharp identified sets.*

In practice, to achieve computational feasibility we may need to choose a small $\bar{a}$ (e.g. $\bar{a} = 2$ or 3) if n is large. This can lead to information loss due to the dependence of links in a subnetwork. Although the links in a subnetwork have diminishing spillover effects on each other as n increases, their dependence is persistent because of the interaction with the remainder of the network, through both the pairwise stability of the remainder and the equilibrium selection mechanism. The Aldous-Hoover representation in (15) shows that under exchangeability the dependence of the links can be captured by the random variables ξ_0 and $(\xi_i)_{i\geq 1}$ asymptotically. Hence, considering bounds from larger subnetworks is analogous to employing information from restrictions on higher-order moments of functions of these random variables. The magnitude of the information loss in choosing a small $\bar{a}$ then depends on to what extent the information in the joint distribution of these functions can be captured by their lower-order moments.

5 Estimation

In this section, we discuss the estimation of the identified set Θ_I. This set is defined by the conditional moment inequalities

$$\begin{aligned} \mathbb{E}\left[1\{G_{n,a} = g_a\} - H_{1n}(g_a, X_{n,a}, X_{n,-a}; \theta) | X_{n,a}, X_{n,-a}\right] &\leq 0 \\ \mathbb{E}\left[H_{2n}(g_a, X_{n,a}, X_{n,-a}; \theta) - 1\{G_{n,a} = g_a\} | X_{n,a}, X_{n,-a}\right] &\leq 0 \end{aligned} \quad (18)$$

[19] This is because if a parameter value θ satisfies (10) for subnetworks of size a_2 under some equilibrium selection mechanism, then θ also satisfies (10) for subnetworks of size $a_1 < a_2$ under that equilibrium selection mechanism, by adding up all possible constellations of the links in $[a_2] \setminus [a_1]$.

[20] Note that the identified sets in (17) do not necessarily decrease in a because the "nonsharp relaxation" in the bounds tends to be larger as a increases. For example, if there is a region of ε_n such that subnetworks $g_3 = (1, 1, 1)$ and $g_3 = (1, 0, 0)$ are multiple equilibria, then this region is included in the upper bounds of both subnetworks. However, the upper bound for the subnetwork $g_2 = 1$ counts this region only once, so it is not necessarily larger than the sum of the two upper bounds for g_3.

for all g_a, all $(X_{n,a}, X_{n,-a})$, and all $a \leq \bar{a}$. Note that the bounds are invariant under permutations over $[a]^c$, so $X_{n,-a}$ can be replaced by its empirical distribution $\phi_n(X_{n,-a}) = \frac{1}{n-a}\sum_{j\notin[a]}\delta_{X_{n,j}}$ without information loss. This substantially reduces the dimension of the conditioning variables and prevents it from increasing in n.

We further transform the conditional moment inequalities into equivalent unconditional moment inequalities of the form

$$\begin{aligned}
\mathbb{E}\left[\left(1\{G_{n,a}=g_a\} - H_{1n}\left(g_a, X_{n,a}, \phi_n(X_{n,-a});\theta\right)\right) q\left(X_{n,a}, \phi_n(X_{n,-a})\right)\right] &\leq 0 \\
\mathbb{E}\left[\left(H_{2n}\left(g_a, X_{n,a}, \phi_n(X_{n,-a});\theta\right) - 1\{G_{n,a}=g_a\}\right) q\left(X_{n,a}, \phi_n(X_{n,-a})\right)\right] &\leq 0 \quad (19)
\end{aligned}$$

for all nonnegative functions $q(X_{n,a}, \phi_n(X_{n,-a})) \in \mathcal{Q}$, where q represents instruments that depend on the conditioning variables and $\mathcal{Q}$ is a collection of instruments. For discrete X, we can choose $\mathcal{Q} = \{1\{X_{n,a}=x_a\}\cdot\frac{1}{n-a}\sum_{j\notin[a]}1\{X_{n,j}=x\} : \forall x_a \in \mathcal{X}_a, \forall x \in \mathcal{X}\}$. If X is continuous, we follow Andrews and Shi (2013) and choose $\mathcal{Q}$ to be a countable set whose elements approximate nonnegative q well, so there is no information loss in the unconditional moment inequalities. For example, we can transform each $X_{n,i}$ to lie in $[0,1]^{d_x}$ and choose $\mathcal{Q}$ to be a collection of indicator functions of cubes in $[0,1]^{d_x}$ with side lengths decreasing to 0, e.g., $\mathcal{Q} = \{1\{X_{n,a}\in C_a\}\cdot\frac{1}{n-a}\sum_{j\notin[a]}1\{X_{n,j}\in C\} : \forall C_a \in \mathcal{C}_a, \forall C \in \mathcal{C}\}$, where $\mathcal{C} = \{\bigotimes_{d=1}^{d_x}(\frac{k_d-1}{2r}, \frac{k_d}{2r}] : 1 \leq k_d \leq 2r, 1 \leq d \leq d_x, r = r_0, r_0+1, \ldots\}$ for some positive integer r_0, and $\mathcal{C}_a = \bigotimes_{i=1}^{a}\mathcal{C}$ (with abuse of notation denote $X_{n,a} = (X_{n,i})_{i\in[a]}$). In practice, if $\mathcal{Q}$ is infinite, we approximate it by a finite set via truncation or simulation. See Andrews and Shi (2013) for more details. Note that given the choice of the instruments the unconditional moment inequalities contain terms of the form $1\{G_{n,a}=g_a\}1\{X_{n,a}=x_a\}$ or $1\{G_{n,a}=g_a\}1\{X_{n,a}\in C_a\}$. These indicator functions are evaluated in the sense of isomorphism. That is, for a given (g_a, x_a) or (g_a, C_a), the "=" and/or "∈" relations hold if they hold for an isomorphism of $(G_{n,a}, X_{n,a})$.

The sample moments can be constructed using subnetworks in any randomly selected subsets of $[n]$ with size a. In particular, let $A_1, A_2, \ldots, A_{N_a}$ be N_a i.i.d. subsets of $[n]$ with size a drawn from the collection of all such subsets, where N_a is a positive integer. We define the sample moments for a network (G_n, X_n) as

$$m_1(\theta; G_n, X_n, g_a, q) = \frac{1}{N_a}\sum_{i=1}^{N_a}\left[\left(1\{G_{n,A_i}=g_a\} - H_{1n}\left(g_a, X_{n,A_i}, \phi_n(X_{n,-A_i});\theta\right)\right)\cdot\right.$$

$$q\left(X_{n,A_i}, \phi_n\left(X_{n,-A_i}\right)\right)]$$

$$m_2\left(\theta; G_n, X_n, g_a, q\right) = \frac{1}{N_a} \sum_{i=1}^{N_a} \left[\left(H_{2n}\left(g_a, X_{n,A_i}, \phi_n\left(X_{n,-A_i}\right); \theta\right) - 1\left\{G_{n,A_i} = g_a\right\}\right) \cdot q\left(X_{n,A_i}, \phi_n\left(X_{n,-A_i}\right)\right)\right] \tag{20}$$

for all $g_a \in \mathcal{G}_a$ and all $q \in \mathcal{Q}$. These are valid moments because by exchangeability $\mathbb{E}m_1\left(\theta; G_n, X_n, g_a, q\right) \leq 0$ and $\mathbb{E}m_2\left(\theta; G_n, X_n, g_a, q\right) \leq 0$ for $\theta \in \Theta_I\left(a\right)$. Moreover, conditional on $\left(G_n, X_n\right)$ the variances of the moments decrease in N_a. Hence by drawing more subnetworks we can reduce the variance of an estimator and improve efficiency.

The estimation and inference of the identified set are a straightforward application of the moment inequality literature. Details are discussed in the appendix.

6 Computation

In this section we discuss how to compute the bounds in (12). Recall that the upper bound is the probability that a subnetwork is PS for some PS complements, and the lower bound has a similar probability form. Computing the events in these probabilities by brute force (e.g., checking all possible complements) is typically infeasible because the number of possible complements is enormous even for a moderate n. We propose a sophisticated method to compute the bounds that is feasible for large n. In the sequel we focus on TU. The case of NTU can be handled similarly but with higher computational costs.[21]

Our idea comes from the fact that the bounds can be equivalently represented as functions of certain maximal and minimal marginal utilities over all PS complements. Because the pairwise stability of a complement can be represented by a set of inequality constraints, the maximal and minimal marginal utilities can be computed by solving constraint optimization problems.

To describe the method precisely, let us introduce some notation. For any $i < j$, denote by $\Delta V_{ij}\left(g_{-ij}, x_{ij}\right)$ the sum of i and j's marginal utilities from link ij that

[21] The computational cost in NTU for n individuals is approximately that in TU for $\sqrt{2}n$ individuals because i and j's proposals for link ij need to be computed separately.

depend on the complement g_{-ij} and attributes x_{ij},

$$\Delta V_{ij}\left(g_{-ij}, x_{ij}\right) = u\left(x_{ij}\right) + u\left(x_{ji}\right) + \frac{1}{n-2}\sum\nolimits_{k\neq i,j}\left(g_{ik}+g_{jk}\right)\gamma_1 + \frac{2}{n-2}\sum\nolimits_{k\neq i,j} g_{ik}g_{jk}\gamma_2.$$

Let $\bar{\varepsilon}_{ij} = \varepsilon_{ij} + \varepsilon_{ji}$ for $i < j$ and $\bar{\varepsilon} = \left(\bar{\varepsilon}_{ij}\right)_{i<j}$. For simplicity we suppress the subscript n in g, x and ε. With abuse of notation we let $\mathcal{PS}\left(x, \bar{\varepsilon}\right)$ denote the collection of PS networks for a given attribute profile x and preference profile $\bar{\varepsilon}$, and let $\mathcal{PS}\left(g_{12}, x, \bar{\varepsilon}_{-12}\right)$ denote the collection of PS complement g_{-12} for a given link g_{12}, attribute profile x, and preference complement profile $\bar{\varepsilon}_{-12} = \left(\bar{\varepsilon}_{ij}\right)_{(i,j)\neq(1,2)}$. Moreover, let $g_{a,-12} = \left(g_{ij}\right)_{i<j\leq a,\ (i,j)\neq(1,2)}$ and $\bar{\varepsilon}_{a,-12} = \left(\bar{\varepsilon}_{ij}\right)_{i<j\leq a,(i,j)\neq(1,2)}$. Note that $g = \left(g_{12}, g_{a,-12}, g_{-a}\right)$ and $\bar{\varepsilon} = \left(\bar{\varepsilon}_{12}, \bar{\varepsilon}_{a,-12}, \bar{\varepsilon}_{-a}\right)$.

We first consider the upper bound

$$H_{1n}\left(g_a, x\right) = \int 1\left\{\exists g_{-a}, \left(g_a, g_{-a}\right) \in \mathcal{PS}\left(x, \bar{\varepsilon}\right)\right\} dF\left(\bar{\varepsilon}\right)$$

It can be represented as

$$H_{1n}\left(g_a, x\right) = \int 1\{\max_{\substack{g_{-a},\ s.t.\\ \left(g_{a,-12}, g_{-a}\right)\in\mathcal{PS}\left(1, x, \bar{\varepsilon}_{-12}\right)}} \Delta V_{12}\left(g_{a,-12}, g_{-a}, x_{12}\right) + \bar{\varepsilon}_{12} \geq 0\} dF\left(\bar{\varepsilon}_{12}, \bar{\varepsilon}_{-12}\right) \tag{21}$$

for all g_a with $g_{12} = 1$, and

$$H_{1n}\left(g_a, x\right) = \int 1\{\min_{\substack{g_{-a},\ s.t.\\ \left(g_{a,-12}, g_{-a}\right)\in\mathcal{PS}\left(0, x, \bar{\varepsilon}_{-12}\right)}} \Delta V_{12}\left(g_{a,-12}, g_{-a}, x_{12}\right) + \bar{\varepsilon}_{12} < 0\} dF\left(\bar{\varepsilon}_{12}, \bar{\varepsilon}_{-12}\right) \tag{22}$$

for all g_a with $g_{12} = 0$, where the maximization and minimization are over g_{-a}. These expressions follow because given any $\bar{\varepsilon}_{-12}$, $\left(1, g_{-12}\right)$ is PS for some g_{-12} if and only if the sum of $\bar{\varepsilon}_{12}$ and the maximal deterministic marginal utility that pair $(1, 2)$ can obtain for any PS g_{-12} is larger than 0, and similarly for $g_{12} = 0$.

Denote the maximum in (21) and minimum in (22) for a given $\bar{\varepsilon}_{-12}$ by $\max \Delta V_{12}(g_a, x, \bar{\varepsilon}_{-12})$ and $\min \Delta V_{12}\left(g_a, x, \bar{\varepsilon}_{-12}\right)$. Let $F_{\bar{\varepsilon}}$ be the CDF of $\bar{\varepsilon}_{ij}$. We can further write the upper bound as

$$H_{1n}\left(g_a, x\right) = \int \left(1 - F_{\bar{\varepsilon}}\left(-\max \Delta V_{12}\left(g_a, x, \bar{\varepsilon}_{-12}\right)\right)\right) dF\left(\bar{\varepsilon}_{-12}\right)$$

for g_a with $g_{12} = 1$ and

$$H_{1n}(g_a, x) = \int F_{\bar{\varepsilon}}\left(-\min \Delta V_{12}\left(g_a, x, \bar{\varepsilon}_{-12}\right)\right) dF\left(\bar{\varepsilon}_{-12}\right)$$

for g_a with $g_{12} = 0$. These expressions indicate that the upper bound can be computed by (i) simulating i.i.d. $\bar{\varepsilon}_{-12}$, (ii) solving the maximization in (21) and minimization in (22) for each simulated $\bar{\varepsilon}_{-12}$, and (iii) taking the averages of the functions $1 - F_{\bar{\varepsilon}}\left(-\max \Delta V_{12}\left(g_a, x, \bar{\varepsilon}_{-12}\right)\right)$ and $F_{\bar{\varepsilon}}\left(-\min \Delta V_{12}\left(g_a, x, \bar{\varepsilon}_{-12}\right)\right)$ over the simulations of $\bar{\varepsilon}_{-12}$.

The complement g_{-a} consists of the edges in g that connect either two nodes outside of $[a]$ or one node in $[a]$ and another outside of $[a]$. The set of edges that connect two nodes outside of $[a]$ form the subnetwork in $[a]^c$, which we denote by $g_{a^c} = (g_{kl})_{a<k<l}$. We call the set of edges that connect one node in $[a]$ and another outside of $[a]$ the *neighborhood* of $[a]$, denoted by $b_a = (g_{ik})_{i \le a, k > a}$. Clearly $g_{-a} = (b_a, g_{a^c})$. While both b_a and g_{a^c} are choice variables in the optimization problems in (21) and (22), their roles are different under the utility specification in (2). In particular, because the marginal utilities of i and j from link ij depend on g_{-ij} only through the neighborhood of the pair $b_{ij} = (g_{ik}, g_{jk})_{k \ne i,j}$, for any $i, j \in [a]$ the marginal utility $\Delta V_{ij}(g_{-ij}, x_{ij})$ depends on g_{-a} only through the neighborhood b_a of $[a]$, i.e., $\Delta V_{ij}(g_{-ij}, x_{ij}) = \Delta V_{ij}(g_{-ij}(g_a, b_a), x_{ij})$. Similarly, for any $k, l \in [a]^c$ the marginal utility $\Delta V_{kl}(g_{-kl}, x_{kl})$ does not depend on the subnetwork g_a, i.e., $\Delta V_{kl}(g_{-kl}, x_{kl}) = \Delta V_{kl}(g_{-kl}(b_a, g_{a^c}), x_{kl})$. Therefore, the optimization problems in (21) and (22) can be written more explicitly as

$$\max_{b_a, g_{c^c}} / \min_{b_a, g_{c^c}} \Delta V_{12}\left(g_{a,-12}, b_a, x_{12}\right) \tag{23}$$

$$s.t.\ g_{ij} = 1\left\{\Delta V_{ij}\left(g_{-ij}\left(g_a, b_a\right), x_{ij}\right) + \bar{\varepsilon}_{ij} \ge 0\right\},\ i < j \le a, (i,j) \ne (1,2) \tag{24}$$

$$g_{ik} = 1\left\{\Delta V_{ik}\left(g_{-ik}\left(g_a, b_a, g_{a^c}\right), x_{ik}\right) + \bar{\varepsilon}_{ik} \ge 0\right\},\ i \le a, k > a \tag{25}$$

$$g_{kl} = 1\left\{\Delta V_{kl}\left(g_{-kl}\left(b_a, g_{a^c}\right), x_{kl}\right) + \bar{\varepsilon}_{kl} \ge 0\right\},\ a < k < l \tag{26}$$

where the inequalities in (24), (25), and (26) ensure that $g_{a,-12}$, b_a, and g_{a^c} are PS, respectively. Note that the subnetwork g_{a^c} does not enter the objective function nor the inequalities in (24). It enters the optimization only through the inequalities in (25) and (26), by affecting the availability of a neighborhood b_a. This feature is important to reduce the complexity of the optimization problems.

In a special case where links are strategic complements, i.e., $\gamma_1 \geq 0$, $\gamma_2 \geq 0$, the network formation game is a supermodular game (Milgrom and Robert, 1990), and any collection of PS networks has the largest and smallest elements, i.e., there exist PS networks g^0 and g^1 such that $g^0 \leq g \leq g^1$ for any PS network g. The largest and smallest PS networks can be computed from the best-response dynamics, where the number of iterations for convergence is no more than $(\#links)^2 \approx n^4/4$ (Topkis, 1979).

For $a = 2$, we can see immediately that the maximum is achieved at the largest PS complement g^1_{-12} and the minimum is achieved at the smallest PS complement g^0_{-12}. Hence the optimization in (23)-(26) amounts to solving for the largest or smallest PS complements for a given g_{12}.

For $a > 2$, there is no guarantee that the maximum and minimum is achieved at the largest and smallest PS complements g^1_{-a} and g^0_{-a} because of the inequality constraints in (24), i.e. the links in $g_{a,-12}$ need to be PS. However, it is easy to show that the maximum can be achieved by replacing the subnetwork g_{a^c} in (25)-(26) with the largest PS subnetwork in $[a]^c$, denoted by $g^1_{a^c}(g_a, b_a)$ and maximizing the objective function over b_a. Similarly, the minimum can be achieved by replacing the subnetwork g_{a^c} in (25)-(26) with the smallest PS subnetwork in $[a]^c$, denoted by $g^0_{a^c}(g_a, b_a)$ and minimizing the objective function over b_a.

In practice, we can implement the maximization/minimization over b_a and the computation of the largest/smallest PS g_{a^c} iteratively. That is, choose an initial b_a, compute the largest/smallest PS g_{a^c} for the initial b_a, solve for the optimal b_a that maximizes/minimizes the objective function under the largest/smallest PS g_{a^c}, update the initial b_a with the optimal b_a, and iterate. This iterative procedure separates the maximization/minimization over b_a from the computation of g_{a^c}, so the maximization/minimization part can be solved using a standard linear integer programming solver, like CPLEX, with the choice variables reduced to b_a whose dimension is only $a \cdot n$.[22] In our simulations, solving such a linear integer programming problem by CPLEX for $n = 100$ and $a = 3$ takes only 0.007 seconds (on a 3.4GHz CPU). More-

[22] The optimization problems in (23)-(26) are not fully linear because of (i) the interaction terms of the form $g_{ik}g_{jk}$ in the marginal utilities and (ii) the indicator restrictions in (24)-(26). Nevertheless, we can apply the linearization techniques in integer programming to fully linearize these problems. In particular, for (i) we can introduce an additional binary variable $y = g_{ik}g_{jk}$ for each $g_{ik}g_{jk}$ with the additional inequalities $y \leq g_{ik}$, $y \leq g_{jk}$, and $y \geq g_{ik} + g_{jk} - 1$. As for (ii), an indicator restriction $g_{ij} = 1\{\Delta V_{ij} + \varepsilon_{ij} \geq 0\}$ is equivalent to the linear inequalities $L(1 - g_{ij}) \leq \Delta V_{ij} + \varepsilon_{ij} < M g_{ij}$ for sufficiently large M and sufficiently small L.

over, because the effect of the links in b_a on the marginal utility of a link in g_{a^c} is at most of the order of $\frac{a}{n-2}$, the iterative procedure is likely to converge fast. In our simulations it typically converges after one iteration.

In the general case without strategic complementarity, we compute PS g_{a^c} by making use of the property of potential games. Recall that in this general case to ensure the existence of a PS network we need to assume TU so that under our utility specification the game can be represented as a potential game. From the property of potential games a PS network is a local maximum of the potential function, so computing PS g_{a^c} amounts to finding local maxima of the potential function. While finding an exact local maximum is a NP problem, it is possible to find an approximate local maximum in polynomial time. For example, Orlin, Punnen and Schulz (2004) show that an ε-local maximum can be found in time polynomial in the problem size and $\frac{1}{\varepsilon}$. Hence, we can solve the optimization problems approximately by replacing the inequalities in (26) with an availability constraint on b_a, i.e., a neighborhood b_a is available if it is PS for some approximate PS g_{a^c}.[23] We expect that the approximation in g_{a^c} has a negligible effect on the optimal value because g_{a^c} plays a role only through the availability of b_a, and the effect of a link on the marginal utility of another is at most at the order of $\frac{1}{n-2}$.

When we solve the optimization problems for $a > 2$, it is possible that for a given $\bar{\varepsilon}_{a,-12}$ a subnetwork $g_{a,-12}$ cannot be PS for any g_{-a} (i.e., the inequalities in (24) can never be satisfied), so the optimization problems have no solution, leading the integrands in (21) and (22) to be zero. This creates nonsmoothness similar as in crude frequency simulators (McFadden (1989), Pakes and Pollard (1989)) which may require a large number of simulations to reduce the simulation error. We follow the GHK algorithm (Hajivassiliou and Ruud (1994), Geweke and Keane (2001)) and propose a smoother simulator so the number of simulations can be smaller. The idea is to simulate $\bar{\varepsilon}_{a,-12}$ and solve the optimization problems in (23)-(26) sequentially for each link in $[a]$. Details about the algorithm can be found in the appendix.

Next we consider the lower bound. It is given by

$$\begin{aligned} & H_{2n}\left(g_a, x\right) \\ = & \int 1\left\{\exists g_{-a}, \left(g_a, g_{-a}\right) \in \mathcal{PS}\left(x, \bar{\varepsilon}\right) \& \forall g_a' \neq g_a, \forall g_{-a}, \left(g_a', g_{-a}\right) \notin \mathcal{PS}\left(x, \bar{\varepsilon}\right)\right\} dF\left(\bar{\varepsilon}\right) \end{aligned}$$

[23] Such problems can be solved using a constraint integer programming solver like SCIP.

$$= \quad 1 - \int 1\left\{\exists g_a' \neq g_a, \exists g_{-a}, \left(g_a', g_{-a}\right) \in \mathcal{PS}\left(x, \bar{\varepsilon}\right)\right\} dF\left(\bar{\varepsilon}\right) \tag{27}$$

where the last equality follows because $\Pr\left(A \cap B^c\right) = \Pr\left(A \cup B\right) - \Pr\left(B\right)$ and the equilibrium set $\mathcal{PS}\left(x, \bar{\varepsilon}\right)$ is nonempty. For $a = 2$, the bounds satisfy $H_{2n}\left(g_{12} = 1, x\right) = 1 - H_{1n}\left(g_{12} = 0, x\right)$ and $H_{2n}\left(g_{12} = 0, x\right) = 1 - H_{1n}\left(g_{12} = 1, x\right)$, so we get the lower bound immediately. For $a > 2$, note that the indicator function in (27) says that there is a subnetwork $g_a' \neq g_a$, with $g_{12}' = 1$ or 0, such that $\left(g_a', g_{-a}\right)$ is PS for some g_{-a}. Therefore, by considering $g_{12}' = 1$ and 0 separately we can represent this indicator function similarly as those in (21) and (22), i.e.,

$$\begin{aligned} & 1\left\{\exists g_a' \neq g_a, \exists g_{-a}, \left(g_a', g_{-a}\right) \in \mathcal{PS}\left(x, \bar{\varepsilon}\right)\right\} \\ = \quad & 1\{ \max_{\substack{g_{a,-12}', g_{-a},\ s.t. \\ \left(g_{a,-12}', g_{-a}\right) \in \mathcal{PS}\left(1, x, \bar{\varepsilon}_{-12}\right),\ \left(1, g_{a,-12}'\right) \neq g_a}} \Delta V_{12}\left(g_{a,-12}', g_{-a}, x_{12}\right) + \bar{\varepsilon}_{12} \geq 0\} \vee \\ & 1\{ \min_{\substack{g_{a,-12}', g_{-a},\ s.t. \\ \left(g_{a,-12}', g_{-a}\right) \in \mathcal{PS}\left(0, x, \bar{\varepsilon}_{-12}\right),\ \left(0, g_{a,-12}'\right) \neq g_a}} \Delta V_{12}\left(g_{a,-12}', g_{-a}, x_{12}\right) + \bar{\varepsilon}_{12} < 0\} \end{aligned} \tag{28}$$

where the maximization/minimization are over $g_{a,-12}'$ and g_{-a}, and $x \vee y = \max\left(x, y\right)$. The event in (28) occurs if $\bar{\varepsilon}_{12}$ is in the union of $[-\max \Delta V_{12}(g_a, x, \bar{\varepsilon}_{-12}), \infty)$ and $(-\infty, -\min \Delta V_{12}(g_a, x, \bar{\varepsilon}_{-12})]$ (one of them may be empty). Hence, the probability that $\bar{\varepsilon}_{12}$ does not lie in this union set for a simulated $\bar{\varepsilon}_{-12}$ gives one simulation of the lower bound H_{2n}.

The optimization problems in (28) can be represented similarly as those in (23)-(26), i.e.,

$$\begin{aligned} \max_{g_{a,-12}', b_a, g_{a^c}} \Big/ \min_{g_{a,-12}', b_a, g_{a^c}} \quad & \Delta V_{12}\left(g_{a,-12}', b_a, x_{12}\right) \\ s.t. \quad & g_{ij}' = 1\left\{\Delta V_{ij}\left(g_{-ij}\left(g_a', b_a\right), x_{ij}\right) + \bar{\varepsilon}_{ij} \geq 0\right\}, \quad i < j \leq a, \ \ (i,j) \neq (1,2) \\ & g_{ik} = 1\left\{\Delta V_{ik}\left(g_{-ik}\left(g_a', b_a, g_{a^c}\right), x_{ik}\right) + \bar{\varepsilon}_{ik} \geq 0\right\}, \quad i \leq a, \ \ k > a \\ & g_{kl} = 1\left\{\Delta V_{kl}\left(g_{-kl}\left(b_a, g_{a^c}\right), x_{kl}\right) + \bar{\varepsilon}_{kl} \geq 0\right\}, \quad a < k < l \\ & g_a' \neq g_a \text{ with } g_{12}' = 1 \text{ (for max) or } 0 \text{ (for min)} \end{aligned}$$

They can be solved using the aforementioned methods.

7 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to evaluate the subnetwork approach developed in the previous sections. We are in particular interested in the performance of the subnetwork bounds in large networks. Throughout the simulations we consider the marginal utility specification

$$\Delta U_{ij}\left(G_{-ij}, X_i, X_j, \varepsilon_{ij}\right) = \beta\left|X_i - X_j\right| + \frac{1}{n-2}\sum_{k\neq i,j} G_{ik}G_{jk}\gamma + \varepsilon_{ij}$$

where X_i, $i = 1, \ldots, n$, are i.i.d. binary variables that equal to 1 or 0 with equal probability, and ε_{ij}, $i, j = 1, \ldots, n$, $i \neq j$, are i.i.d. $N(0,1)$. The parameter of interest is $\theta = (\beta, \gamma)$. We set the true $\beta_0 = -1$ to create homophily, and assume $\gamma \geq 0$, so the links are strategic complements.[24] We consider pairwise stability in TU. To generate a sample of networks, we compute the largest and smallest PS networks in each observation using the best-response dynamics[25] and let half of the networks in the sample be the largest PS networks and another half the smallest PS networks.

We first investigate the properties of the subnetwork bounds as the network size n increases. To do this, we fix the size of subnetworks at $a = 2$ and consider a variety of network sizes $n = 10, 25, 50, 100, 250, 500$. For each n and each $\gamma \in [0,3]$, we compute the upper and lower bounds for the subnetwork choice probabilities $\Pr\left(G_{12} = 1 \middle| X_1 = 1, X_2 = 1\right)$ and $\Pr\left(G_{12} = 1 \middle| X_1 = 0, X_2 = 1\right)$ with 100 simulations. The bounds are plotted in Figure 3.

We can see in Figure 3 that all the upper and lower bounds tend to converge as $n \to \infty$. The changes in the bounds as n increases become negligible when $n \geq 100$ for all γ. The limits that the bounds tend to converge to are also nontrivial. The bounds are close to 1 only for large γ (e.g., $\gamma \geq 2.5$) when the utility externality from friends in common is huge. For such γ, we expect the networks to be complete, so it is reasonable to get close-to-one bounds. The lowest bounds are achieved at $\gamma = 0$ when there is no externality. In this case, the networks coincide with Erdős-Rényi random

[24] We are grateful to the referee who suggested to consider the case of strategic complementarity. In fact, this is the only case that we are able to generate networks with a large number of individuals.

[25] Under strategic complementarity, the best-response dynamics converge to the largest PS network if the initial network is chosen to be the largest possible network (e.g. the complete network) and converge to the smallest PS network if the initial network is the smallest possible network (e.g. the empty network).

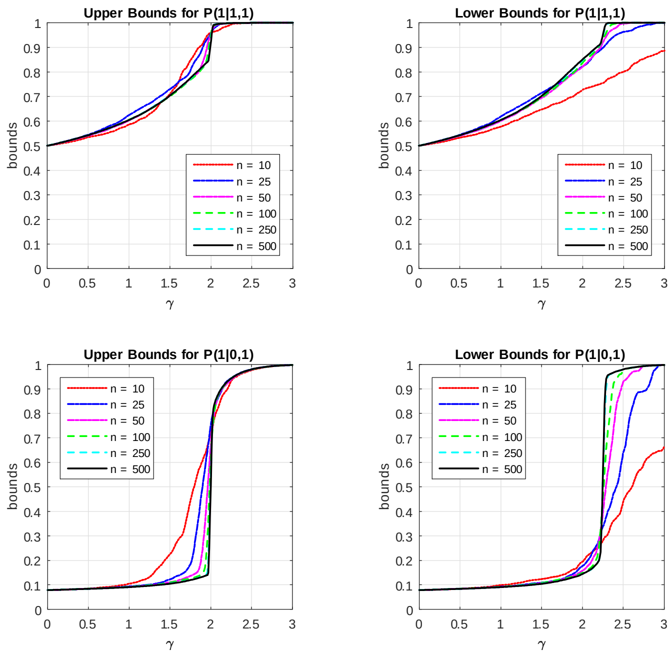

Figure 3: Bounds for Subnetwork Choice Probabilities

graphs with link probability 0.5 for pairs with $X_i = X_j$ and $\Phi\left(-\sqrt{2}\right) = 0.079$ for pairs with $X_i \neq X_j$. The bounds we compute are consistent with these link probabilities. In addition, we also find that the bounds become tighter as n increases, especially for large γ. For example, the difference between the upper and lower bounds for $\Pr\left(G_{12} = 1 \middle| X_1 = 0, X_2 = 1\right)$ at $\gamma = 2.5$ shrinks substantially when n increases from 10 to 50. This finding suggests that the subnetwork bounds may be *more informative* in larger networks,[26] an interesting feature that is worth further research.

Next we examine whether the subnetwork bounds are informative about the parameter. We set the true $\gamma_0 = 1$ and generate i.i.d. networks of sizes $n = 25, 50, 100$ with sample sizes $T = 50, 200$. For each sample, we consider the bounds from subnetworks of sizes $a = 2$ and $a = 3$ and estimate the corresponding identified sets. We compute the bounds using the methods described in Section 6 with 50 simulations, and construct the sample moments in (20) using 1000 random selected subnetworks.

[26]Because the difference between an upper and lower bound reflects the presence of multiple equilibria, its decline implies that multiple equilibria become less prevalent as n increases. This is plausible because intuitively multiple equilibria that differ only in a few number of links may reduce to the "same equilibrium" as $n \to \infty$ due to the averaging in the utility.

For $a = 3$, we also use a graph isomorphism algorithm to determine whether subnetworks are isomorphic.[27] The identified sets are computed using the simulation method suggested by Kline and Tamer (2015). In particular, for an identified set defined as $\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$ for some function $Q \geq 0$, we simulate random variables from a density proportional to $f_{\Theta_I,\rho}(\theta) = \exp\left(-\frac{Q(\theta)}{\rho}\right)$, where $\rho > 0$ is a small tuning parameter,[28] and use the support of the simulated values to approximate the identified set. We implement the simulations by slice sampling (Neal, 2003). Each identified set is approximated by 100 draws. All the aforementioned experiments are repeated independently 200 times.

The estimated identified sets are two-dimensional. For each of them, we calculate its one-dimensional projections, i.e., the maximal and minimal values of the simulated β and γ. Then we pool these maxima and minima from the 200 repetitions of each experiment, and calculate their averages, 5% percentiles of the minima, and 95% percentiles of the maxima. These numbers are reported in Table 1 as the mean estimates and confidence intervals for the one-dimensional projections of the identified sets.Moreover, for each experiment we also calculate the values of θ that are covered by the unions of 90%, 95%, or 99% of the 200 estimated identified sets, and plot them in Figure 4 as the 90%, 95%, and 99% confidence regions of the identified set. Figure 4 is for $T = 50$. The graphs for $T = 200$ are almost identical and thus are omitted.

From Table 1 we can see that the bounds from small subnetworks provide informative estimates for the parameter in all the experiments. In particular, the estimates remain stable when we increase the size of the networks. More interestingly, the confidence intervals for γ tend to be tighter in larger networks. This feature is shown more clearly in Figure 4. The confidence regions of the parameter for $a = 2$ narrow down as n increases from 25 to 100. These results support our earlier findings in the bound experiments and suggest that the subnetwork bounds are informative about the parameter regardless of the size of the networks. The estimation precision for the smallest subnetworks tend to be higher in larger networks.

Moreover, Table 1 shows that bounds from triples ($a = 3$) are more informative than those from pairs ($a = 2$). For example, the upper bounds of β and the lower

[27] We use the graph isomorphism algorithm named Nauty developed by Brendan McKay (http://cs.anu.edu.au/~bdm/nauty). It can calculate isomorphisms for vertex-colored graphs. We transform a subnetwork (g_a, x_a) into a vertex-colored graph, where the colors of the vertices are defined by x_a, so Nauty is applicable.

[28] In the simulations we choose $\rho = 10^{-4}$.

Table 1: Projections of the Estimated Identified Sets

		$a=2$		$a=3$	
T	n	β	γ	β	γ
50	25	$[-1.107, -0.921]$	$[0.813, 1.137]$	$[-1.071, -0.934]$	$[0.868, 1.136]$
		$([-1.205, -0.868])$	$([0.538, 1.338])$	$([-1.115, -0.903])$	$([0.728, 1.251])$
	50	$[-1.107, -0.915]$	$[0.787, 1.129]$	$[-1.069, -0.937]$	$[0.864, 1.123]$
		$([-1.190, -0.860])$	$([0.575, 1.264])$	$([-1.105, -0.912])$	$([0.770, 1.208])$
	100	$[-1.101, -0.917]$	$[0.806, 1.138]$	$[-1.072, -0.937]$	$[0.863, 1.123]$
		$([-1.163, -0.876])$	$([0.621, 1.259])$	$([-1.106, -0.910])$	$([0.772, 1.192])$
200	25	$[-1.104, -0.919]$	$[0.807, 1.126]$	$[-1.071, -0.934]$	$[0.868, 1.132]$
		$([-1.181, -0.866])$	$([0.576, 1.308])$	$([-1.111, -0.908])$	$([0.765, 1.229])$
	50	$[-1.106, -0.915]$	$[0.794, 1.133]$	$[-1.070, -0.937]$	$[0.866, 1.126]$
		$([-1.190, -0.867])$	$([0.565, 1.262])$	$([-1.101, -0.911])$	$([0.771, 1.194])$
	100	$[-1.100, -0.917]$	$[0.808, 1.137]$	$[-1.072, -0.936]$	$[0.859, 1.126]$
		$([-1.162, -0.873])$	$([0.616, 1.261])$	$([-1.107, -0.910])$	$([0.766, 1.198])$
DGP		-1	1	-1	1

Notes: Intervals not in parentheses are the averages of the projections of the identified sets. Intervals in parentheses are the 5% and 95% percentiles of the projections. T is the sample size, n is the network size and a is the subnetwork size.

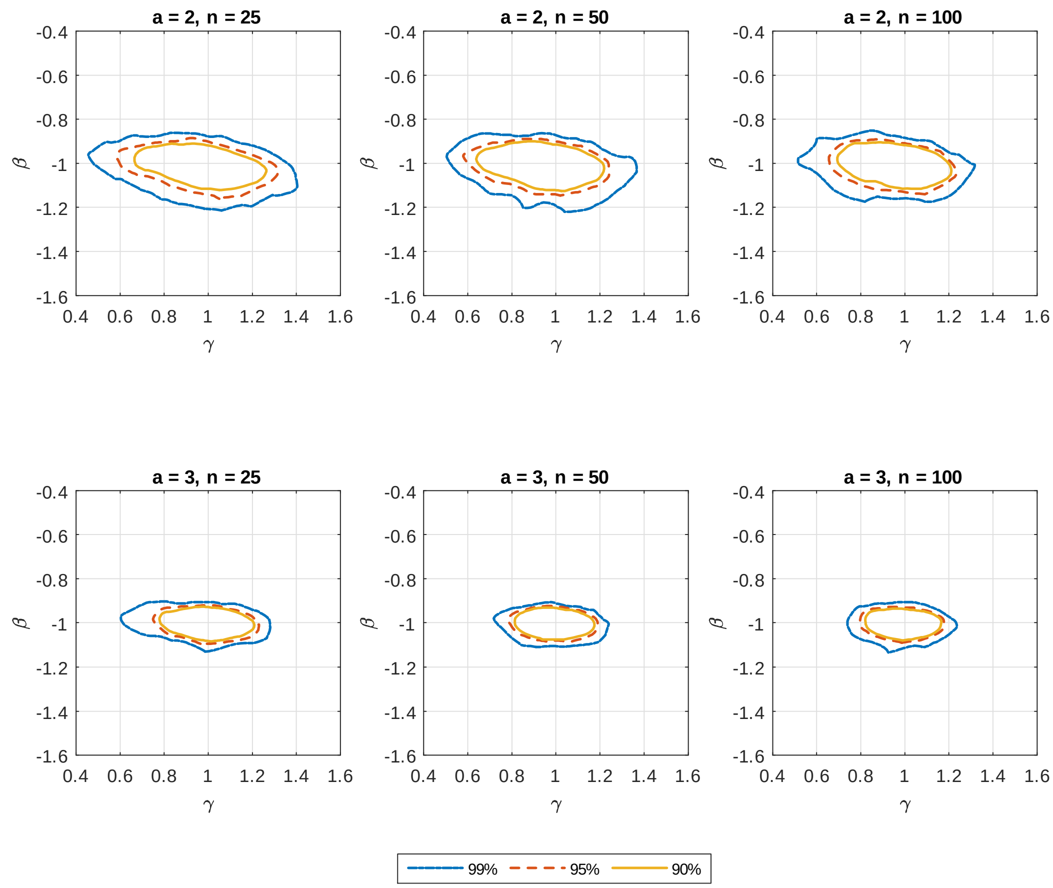

Figure 4: Confidence Regions for the Identified Sets

bounds of γ become tighter in all the mean estimates and confidence intervals when we move from pairs to triples. The same pattern is observed in the confidence regions in Figure 4. These findings suggest that larger subnetworks can provide more information about the parameter, though the improvement seems to be small.

In addition, we find in Table 1 that the estimates in small samples ($T = 50$) are almost identical to those in large samples ($T = 200$). Averaging over a large number of randomly selected subnetworks seems to improve the finite sample performance.

8 Conclusion

In this paper, we develop a structural model of network formation. We characterize network formation as a simultaneous-move game, where the decision of forming a link may depend on the linking decisions of others due to utility externalities from indirect friends. With the prevalence of multiple equilibria, the parameters are not necessarily point identified. We propose a partial identification approach that is computationally feasible in large networks. We derive bounds on the probability of observing a subnetwork. These subnetwork bounds are computationally tractable in large networks provided we consider small subnetworks. We provide both theoretical and Monte Carlo evidence that the bounds from small subnetworks are informative about the parameters in large networks.

This subnetwork approach provides a useful framework for exploring the formation of large networks. By focusing on limited aspects of a network rather than solving the full network at once, we can reduce the dimensionality of the problem and ease the computational burden. For this approach to work, we need small subnetworks to be able to carry the information in large networks, which is the case in our context if networks are exchangeable and convergent. It may be possible to extend our approach to more general networks with these features but are not covered in the present paper. For example, the networks we consider in the paper are dense. It may be of interest to see whether and how our approach can be extended to networks that are sparse. Another interesting extension is to relate our approach to the literature on large networks and investigate under what conditions the inference based on subnetworks from a single large network is asymptotically valid. These extensions are left for future research.

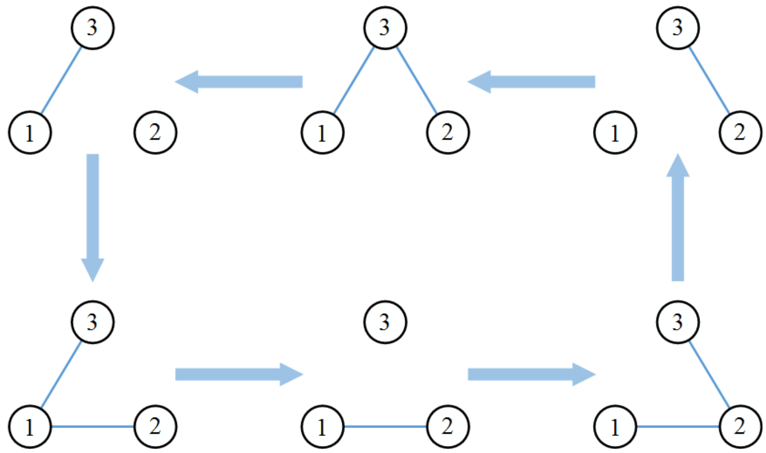

Figure 5: An Example of a Closed Cycle

9 Appendix

9.1 Non-existence of Pairwise Stable Networks

Here we give an example where there is no PS networks, but a closed cycle.

Example 9.1 *Consider networks of size $n = 3$. Suppose the utility function is as in (2) with $u(X_i, X_j; \beta) = 0$, $\gamma_1 < 0$, $\gamma_2 > 0$, $\gamma_1 + \gamma_2 < 0$. Consider the case of NTU. For $\varepsilon_{21}, \varepsilon_{32}, \varepsilon_{13} \geq -\gamma_1$ and $0 \leq \varepsilon_{12}, \varepsilon_{23}, \varepsilon_{31} < -\gamma_1 - \gamma_2$, there is no PS network, but a closed cycle (see Figure 5).*

9.2 Proofs

Proof of Proposition 2.1. By Theorem 1 in Jackson and Watts (2001), if there is a function $\Pi : \mathcal{G} \to \mathbb{R}$ such that for any G, G' that differ by one link, G' defeats G if and only if $\Pi(G') > \Pi(G)$, then there is no cycle and thus no closed cycle.[29] In the case of TU, G' defeating G means that for any $i \neq j$ such that $G'_{ij} \neq G_{ij}$, $U_i(G') + U_j(G') > U_i(G) + U_j(G)$. Hence, the proof is complete if we can find such a Π for the utility function in (2).

We show that

$$\Pi(G) = \sum_{i=1}^{n}\sum_{j=1}^{n} G_{ij}u_{ij} + \frac{1}{2(n-2)} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ k\neq i}}^{n} G_{ij}G_{jk}\gamma_1 + \frac{1}{3(n-2)} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} G_{ij}G_{ik}G_{jk}\gamma_2$$

[29] A *cycle* is a collection of networks that satisfy condition (i) in the definition of closed cycles.

has the desired property, where $u_{ij} = u(X_i, X_j; \beta) + \varepsilon_{ij}$. Consider G and G' which differ by link ij. Assume without loss of generality that $G = (0, G_{-ij})$ and $G' = (1, G_{-ij})$. It suffices to show that $\Pi(G') - \Pi(G) = \Delta U_{ij}(G_{-ij}) + \Delta U_{ji}(G_{-ij})$. By simple algebra

$$\Pi(G') - \Pi(G) = u_{ij} + u_{ji} + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} G_{jk}\gamma_1 + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} G_{ik}\gamma_1 + \frac{2}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} G_{ik}G_{jk}\gamma_2.$$

Moreover, from (3) we have

$$\begin{aligned}
\Delta U_{ij}(G_{-ij}) &= u_{ij} + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} G_{jk}\gamma_1 + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} G_{ik}G_{jk}\gamma_2 \\
\Delta U_{ji}(G_{-ij}) &= u_{ji} + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} G_{ik}\gamma_1 + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} G_{jk}G_{ik}\gamma_2.
\end{aligned}$$

Hence $\Pi(G') - \Pi(G) = \Delta U_{ij}(G_{-ij}) + \Delta U_{ji}(G_{-ij})$. The proof is complete. ∎

Proof of Proposition 2.2. According to Theorem 1 in Hellmann (2012), if a utility function satisfies convexity in one's own links and strategic complementarity, then there is no closed cycle. A utility function U_i satisfies *convexity in one's own links* if for any $j \neq i$ and $G_{-ij}, G'_{-ij} \in \mathcal{G}_{-ij}$ such that $G_{-ij} = G'_{-ij}$ except that $(G_{-ij})_{ik} = 0$ and $\left(G'_{-ij}\right)_{ik} = 1$ for some $k \neq j$, we have $\Delta U_{ij}(G'_{-ij}) \geq \Delta U_{ij}(G_{-ij})$. In other words, if G'_{-ij} differ from G_{-ij} by adding some links that involve i, the marginal utility of i from link ij with these additional links is larger than without. Moreover, U_i satisfies *strategic complementarity* if for any $j \neq i$ and $G_{-ij}, G'_{-ij} \in \mathcal{G}_{-ij}$ such that $G_{-ij} = G'_{-ij}$ except that $(G_{-ij})_{kl} = 0$ and $\left(G'_{-ij}\right)_{kl} = 1$, for some $k, l \neq i$, we have $\Delta U_{ij}(G'_{-ij}) \geq \Delta U_{ij}(G_{-ij})$. In other words, if G'_{-ij} differ from G_{-ij} by adding some links that do not involve i, the marginal utility of i from link ij given these additional links is larger than without. It suffices to verify that the stated utility function satisfies both properties.

The marginal utility (3) is

$$\Delta U_{ij}(G_{-ij}) = u_{ij} + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk}\gamma_1 + \frac{1}{n-2} \sum_{k \neq i,j} G_{ik}G_{jk}\gamma_2.$$

where $u_{ij} = u(X_i, X_j; \beta) + \varepsilon_{ij}$. Since $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$, changing G_{ik} or G_{jk} from 0 to 1 for some $k \neq i, j$ weakly increases $\Delta U_{ij}(G_{-ij})$. Hence both properties are satisfied. The proof is complete. ■

Proof of Theorem 4.1. We first consider a random subset $A' \subseteq [n]$ with size $|A'| = a$. By the definition of subnetwork densities,

$$\begin{aligned}\Pr\left(G_{n,A'} = g_a, X_{n,A'} = x_a \mid X_n\right) &= \mathbb{E}\left[\Pr\left(G_{n,A'} = g_a, X_{n,A'} = x_a \mid G_n, X_n\right) \mid X_n\right] \\ &= \mathbb{E}\left[t_{ind}\left((g_a, x_a), (G_n, X_n)\right) \mid X_n\right]\end{aligned}$$

and

$$\begin{aligned}\Pr\left(G^*_{A'} = g_a, X^*_{A'} = x_a \mid X^*\right) &= \mathbb{E}\left[\Pr\left(G^*_{A'} = g_a, X^*_{A'} = x_a \mid G^*, X^*\right) \mid X^*\right] \\ &= \mathbb{E}\left[t_{ind}\left((g_a, x_a), (G^*, X^*)\right) \mid X^*\right].\end{aligned}$$

For any fixed $m \geq a$, because $t_{ind}\left((g_a, x_a), (G_n, X_n)\right)$ is bounded (by 1), Assumption 3(ii) implies that

$$\mathbb{E}\left[t_{ind}\left((g_a, x_a), (G_n, X_n)\right) \mid X_m\right] \to \mathbb{E}\left[t_{ind}\left((g_a, x_a), (G^*, X^*)\right) \mid X_m\right], \text{ as } n \to \infty.$$

Moreover, define a sequence of random variables Z_m, $m \geq a$, as

$$Z_m = \mathbb{E}\left[\left(t_{ind}\left((g_a, x_a), (G^*, X^*)\right)\right) \mid X_m\right].$$

Because $\mathbb{E}\left[Z_{m+1} \mid X_m\right] = \mathbb{E}\left[\mathbb{E}\left[\left(t_{ind}\left((g_a, x_a), (G^*, X^*)\right)\right) \mid X_{m+1}\right] \mid X_m\right] = Z_m$, the sequence $\{Z_m, \sigma(X_m)\}_{m \geq a}$ is a martingale, so by the martingale convergence theorem we have

$$Z_m = \mathbb{E}\left[t_{ind}\left((g_a, x_a), (G^*, X^*)\right) \mid X_m\right] \xrightarrow{a.s.} \mathbb{E}\left[t_{ind}\left((g_a, x_a), (G^*, X^*)\right) \mid X^*\right], \text{ as } m \to \infty.$$

Hence

$$\begin{aligned} & \left|\Pr\left(G_{n,A'} = g_a, X_{n,A'} = x_a \mid X_n\right) - \Pr\left(G^*_{A'} = g_a, X^*_{A'} = x_a \mid X^*\right)\right| \\ \leq\ & \left|\mathbb{E}\left[t_{ind}\left((g_a, x_a), (G_n, X_n)\right) \mid X_n\right] - \mathbb{E}\left[t_{ind}\left((g_a, x_a), (G^*, X^*)\right) \mid X_n\right]\right| \\ & + \left|\mathbb{E}\left[t_{ind}\left((g_a, x_a), (G^*, X^*)\right) \mid X_n\right] - \mathbb{E}\left[t_{ind}\left((g_a, x_a), (G^*, X^*)\right) \mid X^*\right]\right| \\ \xrightarrow{a.s.}\ & 0 \end{aligned} \tag{29}$$

as $n \to \infty$.

Now we consider the subset $A = [a]$. Note that $\Pr\left(G_{n,a} = g_a | X_{n,a} = x_a | X_n\right) = \Pr\left(G_{n,A'} = g_a, X_{n,A'} = x_a | X_n\right)$ by the exchangeability of (G_n, X_n). Moreover, if $\Pr(X_{n,a} = x_a | X_n) \neq 0$ (it is either 0 or 1), we have $\Pr(G_{n,a} = g_a, X_{n,a} = x_a | X_n) = \Pr(G_{n,a} = g_a | X_{n,a} = x_a, X_{n,-a})$. Similar results hold for the limiting network. Therefore, the convergence result in (29) yields

$$\Pr\left(G_{n,a} = g_a | X_{n,a} = x_a, X_{n,-a}\right) \xrightarrow{a.s.} \Pr\left(G_a^* = g_a | X_a^* = x_a, X_{-a}^*\right), \text{ as } n \to \infty.$$

The proof is complete. ■

Proof of Corollary 4.2. Define random variables Y_n and Y

$$\begin{aligned} Y_n &= \frac{1}{n-2} \sum_{k' \neq i,j} G_{n,ik'} \\ Y &= \mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_k\right) | \xi_0, \xi_i\right] \end{aligned}$$

Since Y_n and Y are bounded, $Y_n \xrightarrow{d} Y$ if for every $r = 1, 2, \ldots$, $\mathbb{E}Y_n^r \to \mathbb{E}Y^r$ as $n \to \infty$.

We start with $r = 1$. The exchangeability of G_n from Assumption 3(i) implies that $\mathbb{E}Y_n = \mathbb{E}G_{n,ik}$ for an arbitrary $k \neq i, j$. Moreover, Theorem 4.1 and exchangeability imply that for any $(g_a, x_a) \in (\mathcal{G}_a, \mathcal{X}_a)$ and any subset $A \subseteq [n]$ with $|A| = a$,

$$\Pr\left(G_{n,A} = g_a | X_{n,A} = x_a\right) \to \Pr\left(G_A^* = g_a | X_A^* = x_a\right), \quad \text{as } n \to \infty \tag{30}$$

by the dominated convergence theorem. Applying (30) with $A = \{i, k\}$ and $g_a = 1$ we have $\mathbb{E}\left[G_{n,ik} | X_{n,ik}\right] \to \mathbb{E}\left[G_{ik}^* | X_{ik}^*\right]$ as $n \to \infty$. By the dominated convergence theorem again we obtain $\mathbb{E}\left[G_{n,ik}\right] \to \mathbb{E}\left[G_{ik}^*\right]$ as $n \to \infty$. The Aldous-Hoover representation in (15) implies that $\mathbb{E}\left[G_{ik}^*\right] = \mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_k\right)\right] = \mathbb{E}\left[Y\right]$. Therefore, $\mathbb{E}Y_n \to \mathbb{E}Y$ as $n \to \infty$.

For $r = 2$, the second moment satisfies $\mathbb{E}Y_n^2 = \mathbb{E}\left(\frac{1}{n-2}\sum_{k' \neq i,j} G_{n,ik'}\right)^2 = \frac{1}{n-2}\mathbb{E}G_{n,ik} + \frac{n-3}{n-2}\mathbb{E}G_{n,ik}G_{n,il}$ for arbitrary $k, l \neq i, j$ with $k \neq l$, where the last equality follows from the exchangeability of G_n. It suffices to show that $\mathbb{E}G_{n,ik}G_{n,il} \to \mathbb{E}Y^2$ as $n \to \infty$. Applying the implication (30) of Theorem 4.1 twice for $A = \{i, k, l\}$ with $g_a = (1, 1, 1)$

and $g_a = (1, 1, 0)$ yields

$$\begin{aligned}
\Pr\left(G_{n,ik} = 1, G_{n,il} = 1, G_{n,kl} = 1 \middle| X_{n,ikl}\right) &\to \Pr\left(G^*_{ik} = 1, G^*_{il} = 1, G^*_{kl} = 1 \middle| X^*_{ikl}\right) \\
\Pr\left(G_{n,ik} = 1, G_{n,il} = 1, G_{n,kl} = 0 \middle| X_{n,ikl}\right) &\to \Pr\left(G^*_{ik} = 1, G^*_{il} = 1, G^*_{kl} = 0 \middle| X^*_{ikl}\right)
\end{aligned}$$

as $n \to \infty$, where for simplicity we denote $X_{n,ikl} = X_{n,\{i,k,l\}}$ and $X^*_{ikl} = X^*_{\{i,k,l\}}$ and the same for similar terms hereafter. Adding up the two convergent probabilities and integrating out $X_{n,ikl}$ and X^*_{ikl} (which are equal) we get $\mathbb{E}G_{n,ik}G_{n,il} \to \mathbb{E}G^*_{ik}G^*_{il}$ as $n \to \infty$. Applying the Aldous-Hoover representation in (15) again gives us

$$\begin{aligned}
\mathbb{E}G^*_{ik}G^*_{il} &= \mathbb{E}\left[1\left\{W\left(\xi_0, \xi_i, \xi_k\right) \geq \xi_{ik}\right\} 1\left\{W\left(\xi_0, \xi_i, \xi_l\right) \geq \xi_{il}\right\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_k\right) \middle| \xi_0, \xi_i\right] \mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_l\right) \middle| \xi_0, \xi_i\right]\right] \\
&= \mathbb{E}Y^2
\end{aligned}$$

Hence, $\mathbb{E}Y_n^2 \to \mathbb{E}Y^2$ as $n \to \infty$.

Now we consider a general $r \in \mathbb{N}$. The rth moment $\mathbb{E}Y_n^r = \left(\frac{1}{n-2}\sum_{k' \neq i,j} G_{n,ik'}\right)^r$ is a sum of all terms of the form

$$\frac{1}{(n-2)^r} \sum_{\substack{1 \leq k_1 \leq \ldots \leq k_s \leq n-2 \\ k_1, \ldots, k_s \neq i,j}} S(r, s)\, \mathbb{E}\left[G_{n,ik_1} G_{n,ik_2} \cdots G_{n,ik_s}\right] \tag{31}$$

where $1 \leq s \leq \min\{r, n-2\}$, $S(r, s)$ is the number of ways to partition r objects into s non-empty subsets, i.e., $S(r, s) = \sum_{a_1, \ldots, a_s > 0, a_1 + \cdots + a_s = r} \frac{r!}{a_1! \cdots a_s!}$, which is a Stirling number of the second kind. By the exchangeability of G_n all the summands in (31) are equal, and the total number of summands is $\binom{n-2}{s}$, so (31) is equal to $\frac{1}{(n-2)^r}\binom{n-2}{s} S(r, s)\, \mathbb{E}G_{n,ik_1}G_{n,ik_2} \cdots G_{n,ik_s}$, where $k_1, k_2, \ldots, k_s$ denote arbitrary s distinct numbers from $[n] \setminus \{i, j\}$. The Stirling numbers of the second kind satisfy the property $\sum_{s=1}^{\min\{r,n-2\}} S(r, s)(n-2)_s = (n-2)^r$,[30] where $(n-2)_s = (n-2) \cdots (n-2-s+1)$, so the coefficient of (31) $\frac{1}{(n-2)^r}\binom{n-2}{s} S(r, s) = \frac{S(r,s)(n-2)_s}{(n-2)^r s!}$ is bounded by 1 and all the coefficients for $s = 1, \ldots, r$ have a sum bounded by 1. If we can show that for any $1 \leq s \leq \min\{r, n-2\}$, $\mathbb{E}\left[G_{n,ik_1} \cdots G_{n,ik_s}\right] \to \mathbb{E}\left[G^*_{ik_1} \cdots G^*_{ik_s}\right]$

[30] This is because both sides of the equation calculate the number of ways to assign r objects to $n-2$ bins.

as $n \to \infty$, then from

$$\begin{aligned}
\mathbb{E}\left[G^*_{ik_1}\cdots G^*_{ik_s}\right] &= \mathbb{E}\left[\prod_{s'=1}^{s} 1\left\{W\left(\xi_0,\xi_i,\xi_{k_{s'}}\right) \geq \xi_{ik_{s'}}\right\}\right] \\
&= \mathbb{E}\left[\prod_{s'=1}^{s} \mathbb{E}\left[W\left(\xi_0,\xi_i,\xi_{k_{s'}}\right)\middle|\,\xi_0,\xi_i\right]\right] \\
&= \mathbb{E}Y^s
\end{aligned}$$

and $\mathbb{E}Y^r = \mathbb{E}\left(\frac{1}{n-2}\sum_{k=1}^{n-2} Y\right)^r = \frac{1}{(n-2)^r}\sum_{s=1}^{\min\{r,n-2\}}\binom{n-2}{s} S(r,s)\,\mathbb{E}Y^s$ we obtain $\mathbb{E}Y_n^r \to \mathbb{E}Y^r$ as $n \to \infty$. To show $\mathbb{E}\left[G_{n,ik_1}\cdots G_{n,ik_s}\right] \to \mathbb{E}\left[G^*_{ik_1}\cdots G^*_{ik_s}\right]$, we apply the implication (30) of Theorem 4.1 for $A = \{i, k_1, \ldots, k_s\}$ with all possible g_a such that $g_{ik_1} = \cdots = g_{ik_s} = 1$. Summing over all such g_a gives

$$\Pr\left(G_{n,ik_1} = 1,\ldots,G_{n,ik_s} = 1\middle|\,X_{n,ik_1\ldots k_s}\right) \to \Pr\left(G^*_{ik_1} = 1,\ldots,G^*_{ik_s} = 1\middle|\,X^*_{ik_1\ldots k_s}\right)$$

as $n \to \infty$. Taking the expectation for both terms the desired result follows.

The second statement of the corollary can be proved similarly. Define random variables Z_n and Z

$$\begin{aligned}
Z_n &= \frac{1}{n-2}\sum_{k'\neq i,j} G_{n,ik'}G_{n,jk'} \\
Z &= \mathbb{E}\left[W\left(\xi_0,\xi_i,\xi_k\right)W\left(\xi_0,\xi_j,\xi_k\right)\middle|\,\xi_0,\xi_i,\xi_j\right]
\end{aligned}$$

Because Z_n and Z are bounded as well, it suffices to show that for every $r = 1, 2, \ldots$, $\mathbb{E}Z_n^r \to \mathbb{E}Z^r$ as $n \to \infty$.

For $r = 1$, the exchangeability of G_n implies $\mathbb{E}Z_n = \mathbb{E}G_{n,ik}G_{n,jk}$ for an arbitrary $k \neq i, j$. Using an argument similar to the above proof for the second moment of Y_n, we can show that $\mathbb{E}G_{n,ik}G_{n,jk} \to \mathbb{E}G^*_{ik}G^*_{jk}$ as $n \to \infty$. The Aldous-Hoover representation implies

$$\begin{aligned}
\mathbb{E}G^*_{ik}G^*_{jk} &= \mathbb{E}\left[1\left\{W\left(\xi_0,\xi_i,\xi_k\right) \geq \xi_{ik}\right\}1\left\{W\left(\xi_0,\xi_j,\xi_k\right) \geq \xi_{jk}\right\}\right] \\
&= \mathbb{E}\left[W\left(\xi_0,\xi_i,\xi_k\right)W\left(\xi_0,\xi_j,\xi_k\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[W\left(\xi_0,\xi_i,\xi_k\right)W\left(\xi_0,\xi_j,\xi_k\right)\middle|\,\xi_0,\xi_i,\xi_j\right]\right] \\
&= \mathbb{E}Z
\end{aligned}$$

Hence, $\mathbb{E}Z_n \to \mathbb{E}Z$ as $n \to \infty$.

For a general $r \in \mathbb{N}$, like $\mathbb{E}Y_n^r$, the rth moment $\mathbb{E}Z_n^r = \left(\frac{1}{n-2}\sum_{k'\neq i,j} G_{n,ik'}G_{n,jk'}\right)^r$ is a sum of all terms of the form

$$\frac{1}{(n-2)^r}\sum_{\substack{1\leq k_1\leq\ldots\leq k_s\leq n-2\\ k_1,\ldots,k_s\neq i,j}} S(r,s)\,\mathbb{E}\left[G_{n,ik_1}G_{n,jk_1}\cdots G_{n,ik_s}G_{n,jk_s}\right] \tag{32}$$

for $1 \leq s \leq \min\{r, n-2\}$. Following the same argument as above, it suffices to show that for any $1 \leq s \leq \min\{r, n-2\}$, $\mathbb{E}\left[G_{n,ik_1}G_{n,jk_1}\cdots G_{n,ik_s}G_{n,jk_s}\right] \to \mathbb{E}\left[G^*_{ik_1}G^*_{jk_1}\cdots G^*_{ik_s}G^*_{jk_s}\right]$ as $n \to \infty$. This follows from (30) with the choice of $A = \{i, j, k_1, \ldots, k_s\}$ and all possible g_a such that $g_{ik_1} = g_{jk_1} = \cdots = g_{ik_s} = g_{jk_s} = 1$. Summing over all such g_a yields

$$\begin{aligned}
&\Pr\left(G_{n,ik_1} = 1, G_{n,jk_1} = 1\ldots, G_{n,ik_s} = 1, G_{n,jk_s} = 1\middle|\, X_{n,ijk_1\ldots k_s}\right)\\
\to\ &\Pr\left(G^*_{ik_1} = 1, G^*_{jk_1} = 1\ldots, G^*_{ik_s} = 1, G^*_{jk_s} = 1\middle|\, X^*_{ijk_1\ldots k_s}\right)
\end{aligned}$$

as $n \to \infty$. By the Aldous-Hoover representation

$$\begin{aligned}
\mathbb{E}\left[G^*_{ik_1}G^*_{jk_1}\cdots G^*_{ik_s}G^*_{jk_s}\right] &= \mathbb{E}\prod_{s'=1}^{s} 1\{W(\xi_0,\xi_i,\xi_{k_{s'}}) \geq \xi_{ik_{s'}}\}1\{W(\xi_0,\xi_j,\xi_{k_{s'}}) \geq \xi_{jk_{s'}}\}\\
&= \mathbb{E}\prod_{s'=1}^{s} W(\xi_0,\xi_i,\xi_{k_{s'}})W(\xi_0,\xi_j,\xi_{k_{s'}})\\
&= \mathbb{E}\prod_{s'=1}^{s} \mathbb{E}[W(\xi_0,\xi_i,\xi_{k_{s'}})W(\xi_0,\xi_j,\xi_{k_{s'}})\,|\xi_0,\xi_i,\xi_j]\\
&= \mathbb{E}Z^s
\end{aligned}$$

Following the same argument as above we have $\mathbb{E}Z_n^r \to \mathbb{E}Z^r$ as $n \to \infty$. The proof is complete. ■

Proof of Remark 4.1. We prove the remark for discrete X. Assumption 3(ii) implies that for any $x_{12} \in \mathcal{X}^2$,

$$\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=i+1}^{n} G_{n,ij}1\{X_{n,ij} = x_{12}\} \xrightarrow{d} \lim_{n\to\infty}\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=i+1}^{n} G^*_{ij}1\{X^*_{ij} = x_{12}\} \tag{33}$$

as $n \to \infty$. From the Aldous-Hoover representation in (15) a limit link G^*_{ij} can be represented as $G_{ij}^* = 1\{W\left(\xi_0, \xi_i, \xi_j\right) \geq \xi_{ij}\}$, where ξ_0, $(\xi_i)_{i\geq 1}$, and $\left(\xi_{ij}\right)_{j>i\geq 1}$ are i.i.d. $U[0,1]$ random variables. Furthermore, define the function $W_x\left(x_{12}, \xi_0, \xi_i, \xi_j\right) = \Pr\left(f_2\left(\xi_0, \xi_i, \xi_j, \xi_{ij}\right) = x_{12} \left| \xi_0, \xi_i, \xi_j\right.\right)$, where f_2 is the component of f in (14) that corresponds to X_{ij}^*. We can represent the random variable $1\{X_{ij}^* = x_{12}\}$ as $1\{W_x\left(x_{12}, \xi_0, \xi_i, \xi_j\right) \geq \eta_{ij}\}$ for some i.i.d. $U(0,1)$ random variables $\left(\eta_{ij}\right)_{j>i\geq 1}$ that are independent of $(\xi_i)_{i\geq 1}$ and $\left(\xi_{ij}\right)_{j>i\geq 1}$.

Given ξ_0, the strong law of large numbers for U-statistics implies

$$\frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} W\left(\xi_0, \xi_i, \xi_j\right) W_x\left(x_{12}, \xi_0, \xi_i, \xi_j\right) \xrightarrow{a.s} \mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_j\right) W_x\left(x_{12}, \xi_0, \xi_i, \xi_j\right) \middle| \xi_0\right] \tag{34}$$

as $n \to \infty$. Moreover, note that $\mathbb{V}(G_{ij}^* 1\{X_{ij}^* = x_{12}\} | W(\xi_0, \xi_i, \xi_j) W_x(x_{12}, \xi_0, \xi_i, \xi_j)) \leq \mathbb{E}[G_{ij}^* 1\{X_{ij}^* = x_{12}\} | W(\xi_0, \xi_i, \xi_j) W_x(x_{12}, \xi_0, \xi_i, \xi_j)]$ and $\sum_{i=1}^n \sum_{j=i+1}^n W(X_{ij}^*, \xi_0, \xi_i, \xi_j)$ $W_2\left(x_{12}, \xi_0, \xi_i, \xi_j\right) \to \infty$ a.s. as $n \to \infty$ because the limit in (34) is positive as $W \not\equiv 0$ and $W_x \not\equiv 0$. Applying Theorem 16 in Caron and Fox (2015) thus yields

$$\frac{\sum_{i=1}^n \sum_{j=i+1}^n G_{ij}^* 1\{X_{ij}^* = x_{12}\}}{\sum_{i=1}^n \sum_{j=i+1}^n W\left(\xi_0, \xi_i, \xi_j\right) W_x\left(x_{12}, \xi_0, \xi_i, \xi_j\right)} \xrightarrow{a.s} 1 \tag{35}$$

as $n \to \infty$. Combining (33)-(35) we then obtain

$$\frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} G_{n,ij} 1\left\{X_{n,ij} = x_{12}\right\} \xrightarrow{d} \mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_j\right) W_x\left(x_{12}, \xi_0, \xi_i, \xi_j\right) \middle| \xi_0\right]$$

as $n \to \infty$. Conditional on ξ_0 the limit is constant, so the statement also holds for convergence in probability. By Slusky's theorem,

$$\frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} G_{n,ij} \xrightarrow{p} \mathbb{E}\left[W\left(\xi_0, \xi_i, \xi_j\right) \middle| \xi_0\right]$$

because $\sum_{x_{12}} 1\left\{X_{n,ij} = x_{12}\right\} = \sum_{x_{12}} W_x\left(x_{12}, \xi_0, \xi_i, \xi_j\right) = 1$. Since ξ_0 is supported on $[0,1]$, we have $\sum_{i=1}^n \sum_{j=i+1}^n G_{n,ij} = O_p\left(n^2\right)$. ■

Proof of Remark 4.2. For continuous X, we replace Definition 4.2 with the definition below.

Definition 9.1 *A sequence of finite exchangeable networks* $\{(G_n, X_n), n \geq 2\}$ *con-*

verge to an infinite exchangeable network $(G^*, X^*) = \left(G^*_{ij}, X^*_{ij}\right)_{i,j\geq 1, i\neq j}$ *if for any* $a \leq n$*, any subnetwork* g_a*, and any Borel subset* $C_a \subseteq \mathcal{X}_a$ *such that* $\Pr\left(X^*_a \in \partial C_a\right) = 0$*, the random variable* $t_{ind}\left(\left(g_a, C_a\right), \left(G_n, X_n\right)\right)$ *converges in distribution to the random variable* $t_{ind}\left(\left(g_a, C_a\right), \left(G^*, X^*\right)\right)$ *as* $n \to \infty$.

Suppose that Assumption 3 is satisfied for the convergence condition defined by Definition 9.1. We verify that Theorem 4.1 and Corollary 4.2 remain true.

We first consider Theorem 4.1. Let (g_a, x_a) be the subnetwork in the statement. Choose $C_a = \{x'_a \in \mathcal{X}_a : \|x_a - x'_a\| < \varepsilon\}$ for some $\varepsilon > 0$ with $\Pr\left(X^*_a \in \partial C_a\right) = 0$. Here the boundary set $\partial C_a = \{x'_a : \|x_a - x'_a\| = \varepsilon\}$. Following the same argument as in the proof of the theorem for discrete X, we can show that for any random subset $A' \subseteq [n]$ with size $|A'| = a$,

$$\begin{aligned} & \left|\Pr\left(G_{n,A'} = g_a, X_{n,A'} \in C_a \middle| X_n\right) - \Pr\left(G^*_{A'} = g_a, X^*_{A'} \in C_a \middle| X^*\right)\right| \\ = & \left|\mathbb{E}\left[t_{ind}\left(\left(g_a, C_a\right), \left(G_n, X_n\right)\right) \middle| X_n\right] - \mathbb{E}\left[t_{ind}\left(\left(g_a, C_a\right), \left(G^*, X^*\right)\right) \middle| X^*\right]\right| \\ \stackrel{a.s.}{\to} & \ 0 \end{aligned} \tag{36}$$

as $n \to \infty$. Let $A = [a]$. Exchangeability implies that $\Pr(G_{n,A'} = g_a, X_{n,A'} \in C_a \,|X_n\,) = \Pr(G_{n,a} = g_a, X_{n,a} \in C_a \,|X_n\,)$ and similar for the limiting network. By the property of conditional expectation and dominated convergence theorem

$$\begin{aligned} \Pr\left(G_{n,a} = g_a, X_{n,a} \in C_a \,|X_n\,\right) &= \mathbb{E}\left[1\left\{X_{n,a} \in C_a\right\} \Pr(G_{n,a} = g_a | \, X_{n,a}, X_{n,-a}) \,|X_n\right] \\ &\to \mathbb{E}\left[1\left\{X_{n,a} = x_a\right\} \Pr(G_{n,a} = g_a | \, X_{n,a}, X_{n,-a}) \,|X_n\right] \\ &= \Pr(G_{n,a} = g_a | \, X_{n,a} = x_a, X_{n,-a}) \end{aligned} \tag{37}$$

as $\varepsilon \to 0$. Similarly

$$\Pr\left(G^*_a = g_a, X^*_a \in C_a \,|X^*\,\right) \to \Pr(G^*_a = g_a | \, X^*_a = x_a, X^*_{-a}) \tag{38}$$

as $\varepsilon \to 0$. Since the distribution of X^*_a has at most countable discontinuous points, we can choose the sequence of $\varepsilon \to 0$ to be such that C_a for each ε in the sequence satisfies $\Pr\left(X^*_a \in \partial C_a\right) = 0$. Then combining (36) together with (37) and (38), we conclude that

$$\Pr(G_{n,a} = g_a | \, X_{n,a} = x_a, X_{n,-a}) \stackrel{a.s.}{\to} \Pr(G^*_a = g_a | \, X^*_a = x_a, X^*_{-a})$$

as $n \to \infty$. Theorem 4.1 is proved for continuous X.

Since condition (30) is satisfied by Theorem 4.1, Corollary 4.2 holds for continuous X without modifying the proof. ■

Proof of Lemma 4.3. Without loss of generality we assume TU, and the case of NTU can be proved similarly. For any $i < j \leq n$, define $v_{ij}(g_{n,-ij})$ to be the marginal utility of i forming a link with j that is due to the utility externality from other links, i.e.,

$$v_{ij}(g_{n,-ij}) = \frac{1}{n-2}\sum_{k\neq i,j} g_{n,jk}\gamma_1 + \frac{1}{n-2}\sum_{k\neq i,j} g_{n,ik}g_{n,jk}\gamma_2.$$

Since both $\frac{1}{n-2}\sum_{k\neq i,j} g_{n,jk}$ and $\frac{1}{n-2}\sum_{k\neq i,j} g_{n,ik}g_{n,jk}$ are bounded between 0 and 1, there exist finite constants v^l and v^l such that $v^l \leq v_{ij}(g_{n,-ij}) \leq v^u$ for all $g_{n,-ij}$. Let $\bar{u}(x_{ij}) = u(x_i, x_j) + u(x_j, x_i)$ and $\bar{\varepsilon}_{ij} = \varepsilon_{ij} + \varepsilon_{ji}$.

The upper bound $H_{1n}(g_a, x_a, X_{n,-a})$ is the probability that there is $g_{n,-a}$ such that $(g_a, g_{n,-a})$ is pairwise stable. By the definition of pairwise stability for such $g_{n,-a}$ the sum of the marginal utilities of i and j from link ij for any $i < j \leq a$ satisfies

$$\Delta U_{ij} + \Delta U_{ji} = \bar{u}(x_{ij}) + v_{ij}(g_{n,-ij}) + v_{ji}(g_{n,-ij}) + \bar{\varepsilon}_{ij} \begin{cases} \geq 0, & \text{if } g_{ij} = 1 \\ < 0, & \text{if } g_{ij} = 0 \end{cases} \tag{39}$$

Since $2v^l \leq v_{ij}(g_{n,-ij}) + v_{ji}(g_{n,-ij}) \leq 2v^u$, the event in (39) implies that $\bar{u}(x_{ij}) + 2v^u + \bar{\varepsilon}_{ij} \geq 0$ if $g_{ij} = 1$ and $\bar{u}(x_{ij}) + 2v^l + \bar{\varepsilon}_{ij} < 0$ if $g_{ij} = 0$. Hence,

$$\begin{aligned} H_{1n}(g_a, x_a, X_{n,-a}) \quad \leq \quad & \prod_{\substack{i<j\leq a \\ g_{ij}=1}} \Pr\left(\bar{u}(x_{ij}) + 2v^u + \bar{\varepsilon}_{ij} \geq 0\right) \\ & \cdot \prod_{\substack{i<j\leq a \\ g_{ij}=0}} \Pr\left(\bar{u}(x_{ij}) + 2v^l + \bar{\varepsilon}_{ij} < 0\right) \end{aligned}$$

Define the right hand side to be $\bar{H}_1(g_a, x_a)$. It is strictly smaller than 1 because v^u and v^l are bounded.

Similarly, the lower bound $H_{2n}(g_a, x_a, X_{n,-a})$ is the probability that there is $g_{n,-a}$ such that $(g_a, g_{n,-a})$ is pairwise stable and only g_a has this property. For such $g_{n,-a}$ the sum of the marginal utilities of i and j from link ij for $i < j \leq a$ also satisfies the event in (39), which holds if $\bar{u}(x_{ij}) + 2v^l + \bar{\varepsilon}_{ij} \geq 0$ if $g_{ij} = 1$ and $\bar{u}(x_{ij}) + 2v^u + \bar{\varepsilon}_{ij} < 0$ if $g_{ij} = 0$. Moreover, when this event occurs there is no $g'_a \neq g_a$ that can satisfy the pairwise stability condition. Therefore,

$$\begin{aligned} H_{2n}\left(g_a, x_a, X_{n,-a}\right) \;\; &\geq \;\; \prod_{\substack{i<j\leq a \\ g_{ij}=1}} \Pr\left(\bar{u}\left(x_{ij}\right) + 2v^l + \bar{\varepsilon}_{ij} \geq 0\right) \\ &\quad \cdot \prod_{\substack{i<j\leq a \\ g_{ij}=0}} \Pr\left(\bar{u}\left(x_{ij}\right) + 2v^u + \bar{\varepsilon}_{ij} < 0\right), \end{aligned}$$

Define the right hand side to be $\bar{H}_2\left(g_a, x_a\right)$. It is strictly greater than 0 because of the boundedness of v^u and v^l. ■

Proof of Theorem 4.4. We start the proof by observing that the bounds can be represented as subnetwork choice probabilities generated under certain extreme equilibrium selection mechanisms. In particular, for any fixed (g_a, x_a), let λ_{1n} and λ_{2n} denote two types of equilibrium selection mechanisms, where λ_{1n} *always* selects a network with the subnetwork g_a and λ_{2n} *never* selects a network with the subnetwork g_a, whenever possible. That is, for any complement $g_{n,-a}$, the equilibrium selection mechanisms λ_{1n} and λ_{2n} satisfy

$$\lambda_{1n}\left(g'_a, g_{n,-a} \middle| \mathcal{PS}\left(\Delta U_n\left(X_n, \varepsilon_n\right)\right)\right) = 0 \text{ for all } g'_a \neq g_a, \text{ if } g_a \in \mathcal{PS}_{[a]}\left(\Delta U_n\left(X_n, \varepsilon_n\right)\right), \tag{40}$$

and

$$\lambda_{2n}\left(g_a, g_{n,-a} \middle| \mathcal{PS}\left(\Delta U_n\left(X_n, \varepsilon_n\right)\right)\right) = 0, \text{ if there is } g'_a \neq g_a,\, g'_a \in \mathcal{PS}_{[a]}\left(\Delta U_n\left(X_n, \varepsilon_n\right)\right). \tag{41}$$

Denote the networks generated under λ_{1n} and λ_{2n} by G_{1n} and G_{2n}, and their subnetworks in $[a]$ and complements by $G_{1n,a}$, $G_{1n,-a}$ and $G_{2n,a}$, $G_{2n,-a}$, respectively. By definition of the bounds, the upper bound is equal to the probability that subnetwork g_a is observed in G_{1n} and the lower bound is the probability that subnetwork g_a is observed in G_{2n}, i.e.,

$$\begin{aligned} H_{1n}\left(g_a, x_a, X_{n,-a}\right) &= \Pr\left(G_{1n,a} = g_a \middle| X_{n,a} = x_a, X_{n,-a}\right) \\ H_{2n}\left(g_a, x_a, X_{n,-a}\right) &= \Pr\left(G_{2n,a} = g_a \middle| X_{n,a} = x_a, X_{n,-a}\right). \end{aligned}$$

Let $\varepsilon_a = \left(\varepsilon_{ij}\right)_{i,j\leq a, i\neq j}$. For a given complement $g_{n,-a}$, let $\mathcal{PS}\left(\Delta U_a\left(g_{n,-a}, x_a, \varepsilon_a\right)\right)$ be a collection of PS subnetwork in $[a]$, where $\Delta U_a\left(g_{n,-a}, x_a, \varepsilon_a\right) = \{\{\Delta U_{ij}(g_{a,-ij}, g_{n,-a}, x_{ij},$

$\varepsilon_{ij})\}_{g_{a,-ij}}\}_{i,j\leq a,i\neq j}$ is the marginal-utility profile of the individuals in $[a]$. We can further derive the upper bound as

$$\begin{aligned}
&H_{1n}\left(g_a, x_a, X_{n,-a}\right)\\
&= \int \sum_{g_{n,-a}} \lambda_{1n}\left(g_a, g_{n,-a} \middle| \mathcal{PS}\left(\Delta U_n\left(x_a, X_{n,-a}, \varepsilon_n\right)\right)\right) dF\left(\varepsilon_n\right)\\
&= \int \sum_{g_{n,-a}} 1\left\{g_a \in \mathcal{PS}\left(\Delta U_a\left(g_{n,-a}, x_a, \varepsilon_a\right)\right)\right\} \Pr\left(G_{1n,-a} = g_{n,-a} \middle| X_{n,a} = x_a, X_{n,-a}, \varepsilon_n\right) dF\left(\varepsilon_n\right)\\
&= \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_a\left(G_{1n,-a}, x_a, \varepsilon_a\right)\right) \middle| X_{n,a} = x_a, X_{n,-a}\right) \qquad (42)
\end{aligned}$$

and similarly for the lower bound

$$\begin{aligned}
&H_{2n}\left(g_a, x_a, X_{n,-a}\right)\\
&= \int \sum_{g_{n,-a}} \lambda_{2n}\left(g_a, g_{n,-a} \middle| \mathcal{PS}\left(\Delta U_n\left(x_a, X_{n,-a}, \varepsilon_n\right)\right)\right) dF\left(\varepsilon_n\right)\\
&= \int \sum_{g_{n,-a}} 1\left\{\{g_a\} = \mathcal{PS}\left(\Delta U_a\left(g_{n,-a}, x_a, \varepsilon_a\right)\right)\right\} \Pr\left(G_{2n,-a} = g_{n,-a} \middle| X_{n,a} = x_a, X_{n,-a}, \varepsilon_n\right) dF\left(\varepsilon_n\right)\\
&= \Pr\left(\{g_a\} = \mathcal{PS}\left(\Delta U_a\left(G_{2n,-a}, x_a, \varepsilon_a\right)\right) \middle| X_{n,a} = x_a, X_{n,-a}\right). \qquad (43)
\end{aligned}$$

The last expressions in (42) and (43) show that the upper bound is the probability that g_a is a PS subnetwork for a random PS complement $G_{1n,-a}$ generated under the equilibrium selection mechanism λ_{1n}, and the lower bound is the probability that g_a is the unique PS subnetwork for a random PS complement $G_{2n,-a}$ generated under the equilibrium selection mechanism λ_{2n}.[31]

We can see from these expressions that the bounds depend on n only through the complements $(G_{1n,-a}, X_{n,-a})$ and $(G_{2n,-a}, X_{n,-a})$. Moreover, the complements play a role only through the marginal utilities

$$\Delta U_{ij}\left(G_{n,-ij}, x_{ij}, \varepsilon_{ij}\right) = u\left(x_{ij}\right) + \frac{1}{n-2}\sum_{k\neq i,j} G_{n,ik}\gamma_1 + \frac{1}{n-2}\sum_{k\neq i,j} G_{n,ik}G_{n,jk}\gamma_2 + \varepsilon_{ij}$$

for $i, j \leq a$, $i \neq j$. Therefore, if the average terms $\frac{1}{n-2}\sum_{k\neq i,j} G_{n,ik}$ and $\frac{1}{n-2}\sum_{k\neq i,j} G_{n,ik}G_{n,jk}$

[31] Note that for a given x_a and ε_a, whether g_a is a PS subnetwork is completely determined by the complements $G_{1n,-a}$ and $G_{2n,-a}$, which are random because of the randomness in $\varepsilon_{n,-a}$ and equilibrium selection mechanisms.

constructed from the complements $G_{1n,-a}$ and $G_{2n,-a}$ converge as $n \to \infty$, we expect that the bounds also converge.

The expressions in (42) and (43) hold for any equilibrium selection mechanisms λ_{1n} and λ_{2n} that satisfy the restrictions in (40) and (41). Hence we have the freedom to choose λ_{1n} and λ_{2n} so that the generated complements $G_{1n,-a}$ and $G_{2n,-a}$ have the desired properties, thereby yielding the convergence of the bounds. We restrict the λ_{1n} and λ_{2n} similarly to what Assumption 3 imposes on the equilibrium selection mechanism λ_n in data.

First, we choose λ_{1n} and λ_{2n} that do not depend on the labels in $[a]^c = [n] \setminus [a]$, as specified in the condition (13), so the complements $(G_{1n,-a}, X_{n,-a})$ and $(G_{2n,-a}, X_{n,-a})$ are exchangeable in $[a]^c$, i.e., their distributions are invariant under the permutations over $[a]^c$.[32]

Second, we choose two sequences of equilibrium selection mechanisms $\{\lambda_{1n}, n \geq 2\}$ and $\{\lambda_{2n},\, n \geq 2\}$ such that $(G_{1n,-a}, X_{n,-a})$ and $(G_{2n,-a}, X_{n,-a})$ converge to some infinite arrays $\left(G^*_{1,-a}, X^*_{1,-a}\right) = \left(G^*_{1,ij}, X^*_{1,ij}\right)_{i>a \cup j>a, i\neq j}$ and $\left(G^*_{2,-a}, X^*_{2,-a}\right) = \left(G^*_{2,ij}, X^*_{2,ij}\right)_{i>a \cup j>a, i\neq j}$ that are exchangeable in $\mathbb{N} \setminus [a]$, in the sense of empirical distribution convergence.

To be precise, let $A = [a]$ and define a *neighboring vector* of A by $(g_{Ak}, x_{Ak}) = (g_{ik}, x_{ik})_{i \leq a} \in \{0,1\}^a \times \mathcal{X}^a$ for some $k \notin A$, i.e., the vector of links and attributes between individual k and the individuals in A. For any value of a neighboring vector $(g_{A(a+1)},\, x_{A(a+1)})$, define the empirical distribution of neighboring vectors in a finite complement $(G_{n,-a}, X_{n,-a})$ by

$$\mu\left(\left(g_{A(a+1)}, x_{A(a+1)}\right), (G_{n,-a}, X_{n,-a})\right) = \frac{1}{n-a} \sum_{k=a+1}^{n} 1\left\{G_{n,Ak} \leq g_{A(a+1)}, X_{n,Ak} \leq x_{A(a+1)}\right\}$$

and define the limiting empirical distribution of neighboring vectors in an infinite complement $(G_{-a}, X_{-a}) = (G_{ij}, X_{ij})_{i>a \cup j>a, i\neq j}$ by

$$\mu\left(\left(g_{A(a+1)}, x_{A(a+1)}\right), (G_{-a}, X_{-a})\right) = \lim_{n\to\infty} \frac{1}{n-a} \sum_{k=a+1}^{n} 1\left\{G_{Ak} \leq g_{A(a+1)}, X_{Ak} \leq x_{A(a+1)}\right\}$$

We say that a finite complement $(G_{n,-a}, X_{n,-a})$ converges to an infinite complement (G_{-a}, X_{-a}) if the empirical distribution of neighboring vectors in $(G_{n,-a}, X_{n,-a})$ con-

[32] Note that the networks G_{1n} and G_{2n} generated under λ_{1n} and λ_{2n} cannot be exchangeable over $[n]$, because of the labeling-dependent restrictions in (40) and (41).

verges in distribution to the limiting distribution of neighboring vectors in (G_{-a}, X_{-a}), i.e.,

$$\mu\left(\left(g_{A(a+1)}, x_{A(a+1)}\right), (G_{n,-a}, X_{n,-a})\right) \xrightarrow{d} \mu\left(\left(g_{A(a+1)}, x_{A(a+1)}\right), (G_{-a}, X_{-a})\right) \tag{44}$$

as $n \to \infty$. This definition is motivated by the convergence of exchangeable sequences (Chapter 3 in Kallenberg (2005), Theorem 3.2) and is weaker than the convergence of exchangeable arrays in Definition 4.2. We choose λ_{1n} and λ_{2n} such that $(G_{1n,-a}, X_{n,-a})$ and $(G_{2n,-a}, X_{n,-a})$ converge to $\left(G^*_{1,-a}, X^*_{1,-a}\right)$ and $\left(G^*_{2,-a}, X^*_{2,-a}\right)$, respectively, in the sense of neighboring vector distribution convergence. Note that the infinite exchangeable X^* in the statement of the theorem satisfies the convergence condition, so its restriction on $\mathbb{N}\setminus[a]$ gives the limiting $X^*_{1,-a}$ and $X^*_{2,-a}$. We denote both $X^*_{1,-a}$ and $X^*_{2,-a}$ by X^*_{-a}.

Since the infinite complements $\left(G^*_{1,-a}, X^*_{-a}\right)$ and $\left(G^*_{2,-a}, X^*_{-a}\right)$ are exchangeable in $\mathbb{N}\setminus[a]$, their neighboring vectors have the functional representation of de Finetti Theorem (Lemma 7.1 in Kallenberg (2005)), i.e.,

$$\begin{aligned} \left(G^*_{1,Ak}, X^*_{Ak}\right) &= f_1(\xi_{10}, \xi_{1k}) \text{ a.s.,} \quad k = a+1, a+2, \ldots \\ \left(G^*_{2,Ak}, X^*_{Ak}\right) &= f_2(\xi_{20}, \xi_{2k}) \text{ a.s.,} \quad k = a+1, a+2, \ldots \end{aligned} \tag{45}$$

for measurable functions $f_1, f_2 : [0,1]^2 \to \{0,1\}^a \times \mathcal{X}^a$, some i.i.d. $U(0,1)$ random variables ξ_{10} and $(\xi_{1k})_{k \geq a+1}$, and some i.i.d. $U(0,1)$ random variables ξ_{20} and $(\xi_{2k})_{k \geq a+1}$. For each $i \leq a$, we define the functions

$$\begin{aligned} W_{1i}(\xi_{10}) &:= \Pr(G^*_{1,i(a+1)} = 1 \,|\xi_{10}) = \Pr\left(f_{1i}\left(\xi_{10}, \xi_{1(a+1)}\right) = 1 \,|\xi_{10}\right) \\ W_{2i}(\xi_{20}) &:= \Pr(G^*_{2,i(a+1)} = 1 \,|\xi_{20}) = \Pr\left(f_{2i}\left(\xi_{20}, \xi_{2(a+1)}\right) = 1 \,|\xi_{20}\right) \end{aligned}$$

where f_{1i} is the ith component of f_1, and f_{2i} is the ith component of f_2.

Now we combine the convergence condition in (44) together with the representations of the limiting complements in (45) to show the convergence of the average terms in the marginal utilities. We consider $G_{1n,-a}$ first. Let $\bar{x} = \sup\{(x,x) : x \in \mathcal{X}\}$ (it can be ∞^2). Fix ξ_{10}. For any $i \leq a$, we have

$$\frac{1}{n-a} \sum_{k=a+1}^{n} G_{1n,ik} = 1 - \frac{1}{n-a} \sum_{k=a+1}^{n} 1\left\{G_{1n,ik} \leq 0, \{G_{1n,i'k} \leq 1\}_{i' \leq a, i' \neq i}, \{X_{n,i'k} \leq \bar{x}\}_{i' \leq a}\right\}$$

$$
\begin{aligned}
&\xrightarrow{d} 1-\lim_{n\to\infty}\frac{1}{n-a}\sum_{k=a+1}^{n}1\left\{G_{1,ik}^{*}\leq 0,\{G_{1,i'k}^{*}\leq 1\}_{i'\leq a,i'\neq i},\{X_{i'k}^{*}\leq \bar{x}\}_{i'\leq a}\right\}\\
&= \lim_{n\to\infty}\frac{1}{n-a}\sum_{k=a+1}^{n}G_{1,ik}^{*}\\
&= W_{1i}\left(\xi_{10}\right)+o_{p}\left(1\right)
\end{aligned}
$$

as $n\to\infty$ by the convergence condition in (44) and the weak law of large numbers (note that conditional on ξ_{10} the representation in (45) implies that $G_{1,ik}^{*}$ for $k>a$ are i.i.d.). Hence

$$
\frac{1}{n-a}\sum_{k=a+1}^{n}G_{1n,ik}\xrightarrow{d}W_{1i}\left(\xi_{10}\right)
$$

as $n\to\infty$. As for the averages of friends in common, for any $i,j\leq a$, $i\neq j$, we can write

$$
\begin{aligned}
&\frac{1}{n-a}\sum_{k=a+1}^{n}G_{1n,ik}G_{1n,jk}\\
=\ & 1-\frac{1}{n-a}\sum_{k=a+1}^{n}1\left\{G_{1n,ik}\leq 0,\{G_{1n,i'k}\leq 1\}_{i'\leq a,i'\neq i},\{X_{n,i'k}\leq \bar{x}\}_{i'\leq a}\right\}\\
&-\frac{1}{n-a}\sum_{k=a+1}^{n}1\left\{G_{1n,jk}\leq 0,\{G_{1n,i'k}\leq 1\}_{i'\leq a,i'\neq j},\{X_{n,i'k}\leq \bar{x}\}_{i'\leq a}\right\}\\
&+\frac{1}{n-a}\sum_{k=a+1}^{n}1\left\{G_{1n,ik}\leq 0,G_{1n,jk}\leq 0,\{G_{1n,i'k}\leq 1\}_{i'\leq a,i'\neq i,j},\{X_{n,i'k}\leq \bar{x}\}_{i'\leq a}\right\}.
\end{aligned}
$$

The first and second averages in the last expression converge in distribution to $1-W_{1i}\left(\xi_{10}\right)$ and $1-W_{1j}\left(\xi_{10}\right)$, respectively, as $n\to\infty$. As for the third, we can show that

$$
\begin{aligned}
&\frac{1}{n-a}\sum_{k=a+1}^{n}1\left\{G_{1n,ik}\leq 0,G_{1n,jk}\leq 0,\{G_{1n,i'k}\leq 1\}_{i'\leq a,i'\neq i,j},\{X_{n,i'k}\leq \bar{x}\}_{i'\leq a}\right\}\\
\xrightarrow{d}\ & \lim_{n\to\infty}\frac{1}{n-a}\sum_{k=a+1}^{n}1\left\{G_{1,ik}^{*}\leq 0,G_{1,jk}^{*}\leq 0,\left\{G_{1,i'k}^{*}\leq 1\right\}_{i'\leq a,i'\neq i,j},\{X_{i'k}^{*}\leq \bar{x}\}_{i'\leq a}\right\}\\
=\ & \lim_{n\to\infty}\frac{1}{n-a}\sum_{k=a+1}^{n}1\left\{G_{1,ik}^{*}=0,G_{1,jk}^{*}=0\right\}\\
=\ & \left(1-W_{1i}\left(\xi_{10}\right)\right)\left(1-W_{1j}\left(\xi_{10}\right)\right)+o_{p}\left(1\right)
\end{aligned}
$$

as $n \to \infty$, again by the condition (44) and the weak law of large numbers. Then by the Slutsky's theorem, we have that conditional on ξ_0

$$\frac{1}{n-a} \sum_{k=a+1}^{n} G_{1n,ik} G_{1n,jk} \xrightarrow{d} W_{1i}(\xi_{10}) W_{1j}(\xi_{10})$$

as $n \to \infty$.

Similar convergence results hold for $G_{2n,-a}$. Conditional on ξ_{20}, we can show that for any $i, j \leq a$, $i \neq j$,

$$\begin{aligned} \frac{1}{n-a} \sum_{k=a+1}^{n} G_{2n,ik} &\xrightarrow{d} W_{2i}(\xi_{20}) \\ \frac{1}{n-a} \sum_{k=a+1}^{n} G_{2n,ik} G_{2n,jk} &\xrightarrow{d} W_{2i}(\xi_{20}) W_{2j}(\xi_{20}) \end{aligned}$$

as $n \to \infty$.

Once the average terms converge, the marginal utilities converge as well. To see this, for any $i, j \leq a$, $i \neq j$, consider the marginal utility of i from link ij given a complement $G_{n,-ij}$

$$\begin{aligned} \Delta U_{ij}(G_{n,-ij}, x_{ij}, \varepsilon_{ij}) &= u(x_{ij}) + \frac{1}{n-2} \sum_{k \neq i,j} G_{n,ik} \gamma_1 + \frac{1}{n-2} \sum_{k \neq i,j} G_{n,ik} G_{n,jk} \gamma_2 + \varepsilon_{ij} \\ &= u(x_{ij}) + \frac{n-a}{n-2} \left(\frac{1}{n-a} \sum_{k=a+1}^{n} G_{n,ik} \gamma_1 + \frac{1}{n-a} \sum_{k=a+1}^{n} G_{n,ik} G_{n,jk} \gamma_2 \right) \\ &\quad + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{a} G_{n,ik} \gamma_1 + \frac{1}{n-2} \sum_{\substack{k=1 \\ k \neq i,j}}^{a} G_{n,ik} G_{n,jk} \gamma_2 + \varepsilon_{ij} \end{aligned}$$

Note that the last two sum terms are $o_p(1)$ as $n \to \infty$. We apply the above results to the complements $G_{1n,-a}$ and $G_{2n,-a}$ and obtain that conditional on ξ_{10}

$$\begin{aligned} \Delta U_{ij}(g_{a,-ij}, G_{1n,-a}, x_{ij}, \varepsilon_{ij}) &\xrightarrow{d} u(x_{ij}) + W_{1i}(\xi_{10}) \gamma_1 + W_{1i}(\xi_{10}) W_{1j}(\xi_{10}) \gamma_2 + \varepsilon_{ij} \\ &=: \Delta U^*_{1ij}(\xi_{10}, x_{ij}, \varepsilon_{ij}) \end{aligned}$$

and, conditional on ξ_{20}

$$\Delta U_{ij}(g_{a,-ij}, G_{2n,-a}, x_{ij}, \varepsilon_{ij}) \xrightarrow{d} u(x_{ij}) + W_{2i}(\xi_{20}) \gamma_1 + W_{2i}(\xi_{20}) W_{2j}(\xi_{20}) \gamma_2 + \varepsilon_{ij}$$

$$=: \quad \Delta U_{2ij}^* (\xi_{20}, x_{ij}, \varepsilon_{ij})$$

as $n \to \infty$.

We are ready to show the convergence of the bounds. Without loss of generality we assume TU, and the case of NTU can be proved similarly. We start with the upper bound. By the definition of pairwise stability, the event in the upper bound in (42) is given by

$$\begin{aligned}
& 1 \{g_a \in \mathcal{PS} (\Delta U_a (G_{1n,-a}, x_a, \varepsilon_a))\} \\
= & \prod_{\substack{i,j \le a, i \ne j \\ g_{ij}=1}} 1 \{\Delta U_{ij} (g_{a,-ij}, G_{1n,-a}, x_{ij}, \varepsilon_{ij}) + \Delta U_{ji} (g_{a,-ij}, G_{1n,-a}, x_{ji}, \varepsilon_{ji}) \ge 0\} \\
& \cdot \prod_{\substack{i,j \le a, i \ne j \\ g_{ij}=0}} 1 \{\Delta U_{ij} (g_{a,-ij}, G_{1n,-a}, x_{ij}, \varepsilon_{ij}) + \Delta U_{ji} (g_{a,-ij}, G_{1n,-a}, x_{ji}, \varepsilon_{ji}) < 0\}
\end{aligned}$$

This event defines a bounded and almost surely continuous function of the marginal utilities, because of the continuity assumption on the distribution of ε in Assumption 1. Therefore, by the Portmanteau theorem we can show that for any fixed $m \ge a$,

$$\begin{aligned}
& \Pr (g_a \in \mathcal{PS} (\Delta U_a (G_{1n,-a}, x_a, \varepsilon_a)) | X_{m,a} = x_a, X_{m,-a}) \\
\to & \mathbb{E} \left[\prod_{\substack{i,j \le a, i \ne j \\ g_{ij}=1}} 1 \left\{ \Delta U_{1ij}^* (\xi_{10}, x_{ij}, \varepsilon_{ij}) + \Delta U_{1ji}^* (\xi_{10}, x_{ji}, \varepsilon_{ji}) \ge 0 \right\} \right. \\
& \left. \cdot \prod_{\substack{i,j \le a, i \ne j \\ g_{ij}=0}} 1 \left\{ \Delta U_{1ij}^* (\xi_{10}, x_{ij}, \varepsilon_{ij}) + \Delta U_{1ji}^* (\xi_{10}, x_{ji}, \varepsilon_{ji}) < 0 \right\} \right| X_{m,a} = x_a, X_{m,-a} \Bigg] \\
=: & \Pr (g_a \in \mathcal{PS} (\Delta U_{1a}^* (\xi_{10}, x_a, \varepsilon_a)) | X_{m,a} = x_a, X_{m,-a})
\end{aligned}$$

as $n \to \infty$, where

$$\Delta U_{1a}^* (\xi_{10}, x_a, \varepsilon_a) := \left\{ \Delta U_{1ij}^* (\xi_{10}, x_{ij}, \varepsilon_{ij}) \right\}_{i,j \le a, i \ne j}$$

is the marginal utility profile in $[a]$ for the limiting complement $\left(G_{1,-a}^*, X_{-a}^*\right)$. Then we apply a martingale argument as in Theorem 4.1 to show the convergence of the last display as $m \to \infty$. Define the random variables

$$Y_m = \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_{m,a} = x_a, X_{m,-a}\right), \;\; m \geq a$$

It is easy to see that $\{Y_m, \sigma(X_m)\}_{m \geq a}$ is a martingale, so from the martingale convergence theorem

$$\begin{aligned} & \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_{m,a} = x_a, X_{m,-a}\right) \\ \overset{a.s.}{\to} \; & \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right) \end{aligned}$$

as $m \to \infty$. Combining the previous two convergence results we derive

$$\begin{aligned} & \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_a\left(G_{1n,-a}, x_a, \varepsilon_a\right)\right) \middle| X_{n,a} = x_a, X_{n,-a}\right) \\ \overset{a.s.}{\to} \; & \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right) \end{aligned}$$

as $n \to \infty$.

Note that the limiting upper bound in the last display may be random if the attributes in a network are correlated. In a special case when the attributes in a network are i.i.d., then from the de Finetti representation in (45) X_{-a}^* and ξ_{10} are independent conditional on X_a^* (this is because for $k, l \notin A$ with $k \neq l$, X_{Ak}^* and X_{Al}^* are independent conditional on X_a^*, so X_{Ak}^* and ξ_{10} are independent conditional on X_a^*). In this case, the limiting upper bound satisfies

$$\begin{aligned} & \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right) \\ = \; & \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a\right) \end{aligned}$$

and reduces to a deterministic function.

Because any equilibrium selection mechanism λ_{1n} that satisfies the restriction (40) will give the same upper bound, they must all converge to the same limit. Hence, the limiting upper bound $\Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right)$ is unique for all the choices of λ_{1n} and is thus well defined. We can define this limiting upper bound to be the function $H_1^*\left(g_a, x_a, X_{-a}^*\right)$,

$$H_1^*\left(g_a, x_a, X_{-a}^*\right) := \Pr\left(g_a \in \mathcal{PS}\left(\Delta U_{1a}^*\left(\xi_{10}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right)$$

Then we have proved the convergence of the upper bound

$$H_{1n}\left(g_a, x_a, X_{n,-a}\right) \stackrel{a.s.}{\rightarrow} H_1^*\left(g_a, x_a, X_{-a}^*\right), \quad \text{as } n \rightarrow \infty.$$

The proof for the convergence of the lower bound is almost the same as that for the upper bound. The only difference is that the event in the lower bound in (43)

$$1\left\{\{g_a\} \in \mathcal{PS}\left(\Delta U_a\left(G_{2n,-a}, x_a, \varepsilon_a\right)\right)\right\}$$

does not have a closed form as the event in the upper bound has. Nevertheless, it still defines a bounded and almost surely function of the marginal utilities, so the argument for the upper bound still applies. In particular, define the limiting marginal utility for $i, j \leq a$, $i \neq j$

$$\Delta U_{2ij}^*\left(\xi_{20}, x_{ij}, \varepsilon_{ij}\right) := u\left(x_{ij}\right) + W_{2i}\left(\xi_{20}\right)\gamma_1 + W_{2i}\left(\xi_{20}\right) W_{2j}\left(\xi_{20}\right)\gamma_2 + \varepsilon_{ij}$$

and the limiting marginal utility profile in $[a]$

$$\Delta U_{2a}^*\left(\xi_{20}, x_a, \varepsilon_a\right) := \left\{\Delta U_{2ij}^*\left(\xi_{20}, x_{ij}, \varepsilon_{ij}\right)\right\}_{i,j \leq a, i \neq j}$$

Following a similar argument we can show that the lower bound converges, i.e.,

$$\begin{aligned} & \Pr\left(\{g_a\} = \mathcal{PS}\left(\Delta U_a\left(G_{2n,-a}, x_a, \varepsilon_a\right)\right) \middle| X_{n,a} = x_a, X_{n,-a}\right) \\ \stackrel{a.s.}{\rightarrow} \; & \Pr\left(\{g_a\} = \mathcal{PS}\left(\Delta U_{2a}^*\left(\xi_{20}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right) \end{aligned}$$

as $n \rightarrow \infty$. The limiting lower bound $\Pr\left(\{g_a\} = \mathcal{PS}\left(\Delta U_{2a}^*\left(\xi_{20}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right)$ is unique for all the choices of λ_{2n} satisfying (41). Define it to be the function $H_2^*\left(g_a, x_a, X_{-a}^*\right)$,

$$H_2^*\left(g_a, x_a, X_{-a}^*\right) := \Pr\left(\{g_a\} = \mathcal{PS}\left(\Delta U_{2a}^*\left(\xi_{20}, x_a, \varepsilon_a\right)\right) \middle| X_a^* = x_a, X_{-a}^*\right).$$

We have proved that

$$H_{2n}\left(g_a, x_a, X_{n,-a}\right) \stackrel{a.s.}{\rightarrow} H_2^*\left(g_a, x_a, X_{-a}^*\right), \quad \text{as } n \rightarrow \infty.$$

The proof is complete. ■

9.3 Estimation and Inference of the Identified Set

In this section we discuss the estimation and inference of the identified set. For each observed network (G_{n_t}, X_{n_t}), $t = 1, \ldots, T$, we have moments $m_1(\theta; G_{n_t}, X_{n_t}, g_a, q)$ and $m_2(\theta; G_{n_t}, X_{n_t}, g_a, q)$ defined as in (20) for all $g_a \in \mathcal{G}_a$, all $q \in \mathcal{Q}$, and all $a = 2, \ldots, \bar{a}$. Stack these moments into one vector

$$m_t(\theta) = \left(m_t^2(\theta)', \ldots, m_t^{\bar{a}}(\theta)'\right)'$$

where

$$\begin{aligned} m_t^a(\theta) &= \left(m_{1t}^a(\theta)', m_{2t}^a(\theta)'\right)', \quad a = 2, \ldots, \bar{a}, \\ m_{jt}^a(\theta) &= \left(m_j(\theta; G_{n_t}, X_{n_t}, g_a, q), \forall g_a \in \mathcal{G}_a, \forall q \in \mathcal{Q}\right)', \quad j = 1, 2. \end{aligned}$$

Then the moment inequalities in (19) can be written as

$$\mathbb{E}m_t(\theta) \leq 0. \tag{46}$$

Following the set inference literature (Chernozhukov, Hong and Tamer (hereafter CHT, 2007), Romano and Shaikh (2010), Andrews and Soares (2010) among others), we estimate the identified set by minimizing the sample analogue of a criterion function based on (46). We use the criterion function as in CHT (2007), $Q(\theta) = \left\|(\mathbb{E}m_t(\theta))_+\right\|^2$, where $(x)_+ = \max(x, 0)$ and $\|\cdot\|$ is the Euclidean norm. The identified set is given by

$$\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}. \tag{47}$$

Let $Q_T(\theta) = \left\|(\mathbb{E}_T m_t(\theta))_+\right\|^2$ be the sample analogue of $Q(\theta)$, where $\mathbb{E}_T m_t(\theta) = \frac{1}{T}\sum_{t=1}^T m_t(\theta)$. In practice we use the normalized sample criterion $Q_T'(\theta) = Q_T(\theta) - \inf_{\theta' \in \Theta} Q_T(\theta')$ to account for misspecification (CHT (2007), Ciliberto and Tamer (2009)) and propose the estimator

$$\hat{\Theta}_I = \{\theta \in \Theta : TQ_T'(\theta) \leq c_T\}, \tag{48}$$

where c_T is chosen to be $c_T \to \infty$ and $c_T/T \to 0$, e.g. $c_T \propto \ln T$. Let $d(A, B) = \max\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\}$ be the Hausdorff distance between sets A and B, where $d(a, B) = \inf_{b \in B}\|a - b\|$. It can be shown that $\hat{\Theta}_I$ is consistent under

Hausdorff distance, i.e., $d\left(\hat{\Theta}_I, \Theta_I\right) \xrightarrow{p} 0$ as $T \to \infty$.

Theorem 9.1 *Suppose that Θ is compact and Assumptions 1-3 are satisfied. Then $\hat{\Theta}_I$ in (48) is a consistent estimator of Θ_I, i.e.,*

$$d\left(\hat{\Theta}_I, \Theta_I\right) \xrightarrow{p} 0, \qquad \text{as } T \to \infty.$$

Proof. This is an application of Theorems 3.1 and 4.2 in CHT (2007). We first show that the result follows if $Q(\theta)$ and $Q_T(\theta)$ satisfy (i) $Q(\theta)$ is continuous in θ, (ii) $\sup_{\theta\in\Theta} |Q(\theta) - Q_T(\theta)| = O_p\left(1/\sqrt{T}\right)$, and (iii) $\sup_{\theta\in\Theta_I} Q_T(\theta) = O_p(1/T)$.

To see this, note that condition (iii) implies that $\inf_{\theta\in\Theta} TQ_T(\theta) \leq \inf_{\theta\in\Theta_I} TQ_T(\theta) \leq \sup_{\theta\in\Theta_I} TQ_T(\theta) = O_p(1)$. Hence, from conditions (ii) and (iii) we obtain

$$\sup_{\theta\in\Theta} \sqrt{T} \left|Q(\theta) - Q'_T(\theta)\right| \leq \sup_{\theta\in\Theta} \sqrt{T} \left|Q(\theta) - Q_T(\theta)\right| + \inf_{\theta'\in\Theta} \sqrt{T} Q_T(\theta') = O_p(1),$$

and

$$\sup_{\theta\in\Theta_I} TQ'_T(\theta) = \sup_{\theta\in\Theta_I} TQ_T(\theta) - \inf_{\theta'\in\Theta} TQ_T(\theta') = O_p(1).$$

Following the proof of Theorem 3.1 in CHT, we can show that

$$\sup_{\theta\in\Theta_I} d\left(\theta, \hat{\Theta}_I\right) \xrightarrow{p} 0, \text{ and } \sup_{\theta\in\hat{\Theta}_I} d(\theta, \Theta_I) \xrightarrow{p} 0, \tag{49}$$

which imply $d\left(\hat{\Theta}_I, \Theta_I\right) \xrightarrow{p} 0$. The first part of (49) holds because by choosing $c_T \to \infty$, $\sup_{\theta\in\Theta_I} TQ'_T(\theta) = O_p(1) < c_T$ and thus $\Theta_I \subseteq \hat{\Theta}_I$ with probability approaching one. The second part of (49) follows from

$$\sup_{\theta\in\hat{\Theta}_I} Q(\theta) \leq \sup_{\theta\in\hat{\Theta}_I} \left|Q(\theta) - Q'_T(\theta)\right| + \sup_{\theta\in\hat{\Theta}_I} Q'_T(\theta) \leq O_p\left(\frac{1}{\sqrt{T}}\right) + \frac{c_T}{T} = o_p(1), \tag{50}$$

and

$$\inf_{\theta\in\Theta\backslash\Theta_I^{\varepsilon}} Q(\theta) \geq \delta(\varepsilon), \text{ for some } \delta(\varepsilon) > 0, \tag{51}$$

where $\Theta_I^{\varepsilon} = \{\theta \in \Theta : d(\theta, \Theta_I) < \varepsilon\}$ for $\varepsilon > 0$. (51) is satisfied because $Q(\theta)$ is continuous in θ by condition (i) and thus $\inf_{\theta\in\Theta\backslash\Theta_I^{\varepsilon}} Q(\theta) = Q(\theta^*) > 0$ for some $\theta^* \in \Theta\backslash\Theta_I^{\varepsilon}$ by compactness of Θ. Combining (50) with (51) yields $\hat{\Theta}_I \cap \Theta\backslash\Theta_I^{\varepsilon} = \emptyset$ with probability approaching 1, so we obtain the second part of (49).

Next we show that conditions (i)-(iii) hold under Assumptions 1-3. Without loss of generality we assume TU; the NTU case can be proved similarly. To show condition (i), note that $Q(\theta)$ is continuous in θ if the bounds in (19) are continuous in θ. From (12) the upper bound is $H_{1n}(g_a, X_n; \theta) = \Pr(\exists g_{n,-a}, (g_a, g_{n,-a}) \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon)$. Let $g_{n,-a}^{(1)}, \ldots, g_{n,-a}^{(K)}$ denote the distinct values in $\mathcal{G}_{n,-a}$. The upper bound can be equivalently represented as

$$
\begin{aligned}
& H_{1n}(g_a, X_n; \theta) \\
= & \sum_{k=1}^{K} \Pr\left(\left(g_a, g_{n,-a}^{(k)}\right) \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon\right) \\
& - \sum_{1 \le k_1 < k_2 \le K} \Pr\left(\left(g_a, g_{n,-a}^{(k_1)}\right) \& \left(g_a, g_{n,-a}^{(k_2)}\right) \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon\right) \\
& + \cdots + (-1)^{K-1} \Pr\left(\left(g_a, g_{n,-a}^{(1)}\right) \& \ldots \& \left(g_a, g_{n,-a}^{(K)}\right) \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n, \theta_e\right).
\end{aligned}
$$

For any $i < j \le n$, define $\bar{u}_{ij}(X_{n,ij}; \beta) = u(X_{n,i}, X_{n,j,}; \beta) + u(X_{n,j}, X_{n,i}; \beta)$, $\bar{\varepsilon}_{n,ij} = \varepsilon_{n,ij} + \varepsilon_{n,ji}$, and $\bar{v}_{ij}^{(k)}(\gamma) = \frac{1}{n-2}\sum_{l \ne i,j}(g_{il}^{(k)} + g_{jl}^{(k)})\gamma_1 + \frac{2}{n-2}\sum_{l \ne i,j} g_{il}^{(k)} g_{,jl}^{(k)}\gamma_2$, where the superscript k indicates that the links are from the network $(g_a, g_{n,-a}^{(k)})$, $k = 1, \ldots, K$. The probability terms in the summations in the last display are of the form

$$
\Pr((g_a, g_{n,-a}^{(k_1)}) \& \ldots \& (g_a, g_{n,-a}^{(k_s)}) \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon)
$$

for some $1 \le k_1 < \cdots < k_s \le K$ and some $1 \le s \le K$. By the definition of pairwise stability and Assumptions 1-2, such a term can be written as

$$
\prod_{i<j\le n} \Pr\left(\bar{\varepsilon}_{n,ij} \in D_{ij}(X_{n,ij}; \theta_u) \middle| X_{n,ij}; \theta_\varepsilon\right) \tag{52}
$$

where $D_{ij}(X_{n,ij}; \theta_u) \subseteq \mathbb{R}$ for $i < j \le n$ is an interval of the form

$$
\begin{aligned}
& [-\bar{u}_{ij}(X_{n,ij}; \beta) - \min_{\substack{1 \le r \le s: \\ g_{n,ij}^{(k_r)} = 1}} \bar{v}_{ij}^{(k_r)}(\gamma), \ \infty), \\
& (-\infty, \ -\bar{u}_{ij}(X_{n,ij}; \beta) - \max_{\substack{1 \le r \le s: \\ g_{n,ij}^{(k_r)} = 0}} \bar{v}_{ij}^{(k_r)}(\gamma)), \text{ or} \\
& [-\bar{u}_{ij}(X_{n,ij}; \beta) - \min_{\substack{1 \le r \le s: \\ g_{n,ij}^{(k_r)} = 1}} \bar{v}_{ij}^{(k_r)}(\gamma), \ -\bar{u}_{ij}(X_{n,ij}; \beta) - \max_{\substack{1 \le r \le s: \\ g_{n,ij}^{(k_r)} = 0}} \bar{v}_{ij}^{(k_r)}(\gamma)) \text{ (may be empty)},
\end{aligned}
$$

where $g_{n,ij}^{(k)}$ is the (i,j) element of $g_n^{(k)} = (g_a, g_{n,-a}^{(k)})$, $k = 1, \ldots, K$. Because $\bar{u}_{ij}(X_{n,ij}; \beta)$ and $\bar{v}_{ij}^{(k)}(\gamma)$, $k = 1, \ldots, K$, are continuous in β and γ, max and min are continuous operators, and $\varepsilon_{n,ij}$ has a continuous distribution, we can show that (52) is continuous in $\theta_u = (\beta, \gamma)$. Moreover, the CDF of $\bar{\varepsilon}_{n,ij}$ is continuous in θ_ε under Assumption 1, so (52) is also continuous in θ_ε. Therefore, the upper bound is continuous in θ.

Now we consider the lower bound. Recall from (27) that the lower bound is given by $H_{2n}(g_a, X_n; \theta) = 1 - \Pr(\exists g_a' \neq g_a, \exists g_{n,-a}, (g_a, g_{n,-a}) \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon)$. Let $g_n^{(1)}, \ldots, g_n^{(L)}$ denote the distinct values in $\mathcal{G}_n$ whose subnetwork in $[a]$ is not g_a. Like the upper bound, the probability term in the lower bound can be written as

$$\begin{aligned}
& \Pr\left(\exists g_a' \neq g_a, \exists g_{n,-a}, (g_a, g_{n,-a}) \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon\right) \\
= & \sum_{l=1}^{L} \Pr\left(g_n^{(l)} \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon\right) \\
& - \sum_{1 \leq l_1 < l_2 \leq L} \Pr\left(g_n^{(l_1)} \& g_n^{(l_2)} \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon\right) \\
& + \cdots + (-1)^{L-1} \Pr\left(g_n^{(1)} \& \ldots \& g_n^{(L)} \in \mathcal{PS}(\Delta U_n(X_n, \varepsilon_n; \theta_u)) | X_n; \theta_\varepsilon\right)
\end{aligned}$$

The probability terms in the summations have a representation similar to (52), so using the same argument we can show that they are continuous in θ. This proves that the lower bound is continuous in θ.

As for condition (ii), it suffices to show that $\{m_t(\theta), \theta \in \Theta\}$ is Donsker because then $\sup_{\theta \in \Theta} \sqrt{T} |Q(\theta) - Q_T(\theta)| = \sup_{\theta \in \Theta} \sqrt{T} \left| \left\| (\mathbb{E} m_t(\theta))_+ \right\|^2 - \left\| (\mathbb{E}_T m_t(\theta))_+ \right\|^2 \right| \leq \sup_{\theta \in \Theta} \left\| \sqrt{T} (\mathbb{E}_T m_t(\theta) - \mathbb{E} m_t(\theta)) \right\| \left(\left\| (\mathbb{E}_T m_t(\theta))_+ \right\| + \left\| (\mathbb{E} m_t(\theta))_+ \right\| \right) = O_p(1)$. The collection of moments $\{m_t(\theta), \theta \in \Theta\}$ is Donsker if (1) $m_t(\theta)$ satisfies the finite-dimensional convergence property and (2) $m_t(\theta)$ is stochastically equicontinuous. The finite-dimensional convergence follows from CLT by Assumption 1. As for (2), note that the terms $1\{G_{n,a} = g_a\}$ and $q(X_{n,a}, \phi_n(X_{n,-a}))$ in the moments do not depend on θ, so it suffices to show that the bounds are stochastically equicontinuous, which follows if the functions in (52) are stochastically equicontinuous.

To show stochastic equicontinuity for the functions in (52), observe that the functions $\bar{u}_{ij}(X_{n,ij}; \beta)$ and $\bar{v}_{ij}^{(k)}(\gamma)$, $k = 1, \ldots, K$, span a finite-dimensional vector space, so they form a VC-subgraph class (Van der Vaart and Wellner (1996), Lemma 2.6.15). Therefore, because the VC-subgraph property is closed under max, min and monotonic transformations (note that the CDF of $\bar{\varepsilon}_{n,ij}$ is monotonic), the terms in

(52) as functions of θ_u also form a VC-subgraph class, and thus are stochastically equicontinuous as they are uniformly bounded (by 1). Moreover, the terms in (52) as functions of θ_ε are Lipschitz countinuous in θ_ε because the CDF of $\bar{\varepsilon}_{n,ij}$ is continuously differentiable in θ_ε by Assumption 1 and Θ is compact. Hence, they as functions of θ_ε are also stochastically equicontinuous. Combining the results we can prove that the terms in (52) as functions of θ are stochastically equicontinuous.

Condition (iii) is a result of $\{m_t(\theta), \theta \in \Theta\}$ being Donsker and $\mathbb{E}m_t(\theta) \leq 0$ for $\theta \in \Theta_I$. To see this, note $TQ_T(\theta) = \left\| (\sqrt{T}(\mathbb{E}_T m_t(\theta) - \mathbb{E}m_t(\theta)) + \sqrt{T}\mathbb{E}m_t(\theta))_+ \right\|^2$. For $\theta \in \Theta_I$, we have $TQ_T(\theta) \xrightarrow{p} 0$ as $T \to \infty$ if $\mathbb{E}m_t(\theta) < 0$, and $TQ_T(\theta) = O_p(1)$ if $\mathbb{E}m_t(\theta) = 0$. Hence condition (iii) is satisfied. The proof is complete. ■

The confidence region for the true θ_0 can be constructed by inverting the acceptance region of a test (e.g. CHT (2007), Andrews and Soares (2010), Andrews and Jia (2012)). We use the confidence region proposed by CHT (2007)

$$\mathcal{C}_T = \{\theta \in \Theta : TQ_T'(\theta) \leq \bar{c}_{1-\alpha}(\theta)\}, \tag{53}$$

where $\bar{c}_{1-\alpha}(\theta) = \min(\hat{c}_{1-\alpha}(\theta), \hat{c}_{1-\alpha})$, $\hat{c}_{1-\alpha}(\theta)$ is a consistent estimator of $c_{1-\alpha}(\theta)$, the $1-\alpha$ quantile of the limiting distribution of $TQ_T'(\theta)$, and $\hat{c}_{1-\alpha}$ is a data-dependent variable that is $O_p(1)$ and larger than $\sup_{\theta \in \Theta_I} c_{1-\alpha}(\theta)$. It can be shown that for any $\theta \in \Theta_I$, $\mathcal{C}_T$ is asymptotically correct, i.e., $\liminf_{T\to\infty} \Pr(\theta \subseteq \mathcal{C}_T) \geq 1-\alpha$. We can use the subsampling method in CHT (2007) to obtain $\bar{c}_{1-\alpha}(\theta)$.

9.4 GHK Algorithm for the Computation of the Bounds

In this section, we discuss how to compute the bounds using a GHK algorithm. The algorithms for the upper and lower bounds are similar, so we focus on the upper bound. Instead of simulating $\bar{\varepsilon}_{a,-12}$ and solving the optimization problem in (23)-(26) once, we simulate the components of $\bar{\varepsilon}_{a,-12}$ sequentially and solve a sequence of optimization problems as in (23)-(26) for each link in $[a]$.

For expositional simplicity, we describe the algorithm in an example of $a = 3$.

Algorithm 1 *For a simulated* $\bar{\varepsilon}_{-a}$*,*

1. For $g_{23} = 1$ *or* 0*, (i) solve the problem*

$$\max_{b_a, g_a^c} / \min_{b_a, g_a^c} \quad \Delta V_{23}(g_{a,-23}, b_a, x_{23})$$

$$s.t. \qquad \textit{inequalities (25)-(26)}$$

respectively, and (ii) generate $\bar{\varepsilon}_{23}$ from the conditional distribution

$$\bar{\varepsilon}_{23} \sim \begin{cases} F_{\bar{\varepsilon}}\left(\bar{\varepsilon}_{ij} \middle| \bar{\varepsilon}_{ij} \geq -\max \Delta V_{23}\left(g_{a,-23}, b_a, x_{23}\right)\right), & \textit{if } g_{23} = 1 \\ F_{\bar{\varepsilon}}\left(\bar{\varepsilon}_{ij} \middle| \bar{\varepsilon}_{ij} < -\min \Delta V_{23}\left(g_{a,-23}, b_a, x_{23}\right)\right), & \textit{if } g_{23} = 0 \end{cases}$$

2. For $g_{13} = 1$ or 0, (i) solve the problem

$$\begin{aligned} \max_{b_a, g_a^c} / \min_{b_a, g_a^c} \quad & \Delta V_{13}\left(g_{a,-13}, b_a, x_{13}\right) \\ s.t. \quad & \textit{inequalities (25)-(26)} \\ & g_{23} = 1\left\{\Delta V_{23}\left(g_{a,-23}, b_a, x_{23}\right) + \bar{\varepsilon}_{23} \geq 0\right\} \end{aligned}$$

respectively, and (ii) generate $\bar{\varepsilon}_{13}$ from the conditional distribution

$$\bar{\varepsilon}_{13} \sim \begin{cases} F_{\bar{\varepsilon}}\left(\bar{\varepsilon}_{ij} \middle| \bar{\varepsilon}_{ij} \geq -\max \Delta V_{13}\left(g_{a,-13}, b_a, x_{13}\right)\right), & \textit{if } g_{13} = 1 \\ F_{\bar{\varepsilon}}\left(\bar{\varepsilon}_{ij} \middle| \bar{\varepsilon}_{ij} < -\min \Delta V_{13}\left(g_{a,-13}, b_a, x_{13}\right)\right), & \textit{if } g_{13} = 0 \end{cases}$$

3. For $g_{12} = 1$ or 0, solve the problem

$$\begin{aligned} \max_{b_a, g_a^c} / \min_{b_a, g_a^c} \quad & \Delta V_{12}\left(g_{a,-12}, b_a, x_{12}\right) \\ s.t. \quad & \textit{inequalities (25)-(26)} \\ & g_{23} = 1\left\{\Delta V_{23}\left(g_{a,-23}, b_a, x_{23}\right) + \bar{\varepsilon}_{23} \geq 0\right\} \\ & g_{13} = 1\left\{\Delta V_{13}\left(g_{a,-13}, b_a, x_{13}\right) + \bar{\varepsilon}_{13} \geq 0\right\} \end{aligned}$$

respectively.

Let

$$P_{ij} = \begin{cases} 1 - F_{\bar{\varepsilon}}\left(-\max \Delta V_{ij}\left(g_{a,-ij}, b_a, x_{ij}\right)\right), & \text{if } g_{ij} = 1 \\ F_{\bar{\varepsilon}}\left(-\min \Delta V_{ij}\left(g_{a,-ij}, b_a, x_{ij}\right)\right), & \text{if } g_{ij} = 0 \end{cases}$$

for $i < j \leq 3$. The value $\prod_{i<j\leq 3} P_{ij}$, as a function of $(\bar{\varepsilon}_{13}, \bar{\varepsilon}_{23}, \bar{\varepsilon}_{-a})$, gives one simulation of the upper bound. Repeat the algorithm independently R times. The average of the R values of $\prod_{i<j\leq 3} P_{ij}$ gives a simulator of the upper bound.

In this algorithm, because $\bar{\varepsilon}_{23}$ is generated from a conditional distribution given that g_{23} is PS for some PS g_{-a}, and $\bar{\varepsilon}_{13}$ is generated from a conditional distribution such that (g_{13}, g_{23}) is PS for some PS g_{-a}, the optimization problems in steps 2-3 for such $(\bar{\varepsilon}_{13}, \bar{\varepsilon}_{23})$ are guaranteed to have a solution, and thus the integrands in (21)-(22) are given by P_{12} (rather than 0). Weighting P_{12} appropriately (i.e. multiplying it by $P_{13}P_{23}$) to account for the difference between the conditional distribution and unconditional distribution of $(\bar{\varepsilon}_{13}, \bar{\varepsilon}_{23})$, we obtain a simulator for the upper bound that is "more continuous" in the parameter than the simulator directly solved from (23)-(26).

References

[1] Andrews, D. W. K., S., Berry, and P. Jia (2004) Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location, working paper.

[2] Andrews, D. W. K., and P. Jia (2012) Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure, *Econometrica*, 80(6): 2805–2826.

[3] Andrews, D. W. K., and X. Shi (2013) Inference Based on Conditinal Moment Inequalities, *Econometrica*, 81(2): 609-666.

[4] Andrews, D. W. K., and G. Soares (2010) Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection, *Econometrica*, 78(1): 119-157.

[5] Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011) A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models, *International Economic Review*, 53 (3), 807-824.

[6] Bajari, P., H. Hong., and S. P. Ryan (2010) Identification and Estimation of a Discrete Game of Complete Information, *Econometrica*, 78(5): 1529-1568.

[7] Belleflamme, P., and F. Bloch (2004) Market Sharing Agreements and Collusive Networks. *International Economic Review*, 45(2): 387-411.

[8] Beresteanu, A., I. Molchanov, and F. Molinari (2011) Sharp Identification Regions in Models With Convex Moment Predictions, *Econometrica*, 79(6): 1785–1821.

[9] Berry, S., and E. Tamer (2006) Identification in Models of Oligopoly Entry, in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Volume II, edited by R. Blundell, W. K. Newey, and T. Persson, Cambridge University Press.

[10] Bhamidi, S., G. Bresler, and A. Sly (2011) Mixing Time of Exponential Random Graphs, *The Annals of Applied Probability*, 21(6): 2146-2170.

[11] Bierlaire, M., D. Bolduc, and D. McFadden (2008) The Estimation of Generalized Extreme Value Models from Choice-based Samples, *Transportation Research Part B: Methodological,* 42(4): 381-394.

[12] Bloch, F., and M. O. Jackson (2006) Definitions of Equilibrium in Network Formation Games, *International Journal of Game Theory*, 34: 305-318.

[13] Bloch, F., and M. O. Jackson (2007) The Formation of Networks with Transfers Among Players, *Journal of Economic Theory* 133: 83-110.

[14] Bresnahan, T. F., and P. C. Reiss (1991) Empirical Models of Discrete Games, *Journal of Econometrics*, 48: 57-81.

[15] Bollobás, B. and O. Riordan (2009) Metrics for Sparse Graphs, arXiv: 0708.1919v3.

[16] Boucher, V., and I. Mourifié (2013) My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks, working paper.

[17] Calvó-Armengol, A., and M. O. Jackson (2004) The Effects of Social Networks on Employment and Inequality, *American Economic Review*, 94(3): 426-454.

[18] Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009) Peer Effects and Social Networks in Education, *Review of Economic Studies*, 76, 1239–1267.

[19] Caron, F. and E. Fox (2015) Sparse Graphs Using Exchangeable Random Meaures, *arXiv: 1401.1137.*

[20] Chandrasekhar, A. G., and M. O. Jackson (2013) Tractable and Consistent Random Graph Models, working paper.

[21] Chernozhukov, V., and H. Hong (2004) Likelihood Estimation and Inference in a Class of Nonregular Econometric Models, *Econometrica*, 72(5), 1445-1480.

[22] Chernozhukov, V., H. Hong, and E. Tamer (2007) Estimation and Confidence Regions for Parameter Sets in Econometric Models, *Econometrica*, 75(5), 1243-1284.

[23] Christakis, N., J. Fowler, G. W. Imbens, and K. Kalyanaraman (2010) An Empirical Model for Strategic Network Formation, NBER working paper No.16039.

[24] Ciliberto, F., and E. Tamer (2009) Market Structure and Multiple Equilibria in Airline Markets, *Econometrica, 77(6)*, 1791-1828.

[25] Conley, T., and C. Udry (2010) Learning About a New Technology: Pineapple in Ghana, *American Economic Review*, 100(1), 35-69.

[26] Currarini S., M. O. Jackson and P. Pin (2009) An Economic Model of Friendship: Homophily, Minorities, and Segregation, *Econometrica*, 77(4): 1003-1045.

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

[28] Diaconis, P. and S. Janson (2008) Graph Limits and Exchangeable Random Graphs, arXiv: 0712.2749v1.

[29] Dutta, B., and S. Mutuswami (1997) Stable Networks, *Journal of Economic Theory* 76: 322-344.

[30] Erdős, P., and A. Rényi (1959) On Random Graphs I, *Publicationes Mathematicae Debrecen* 6: 290–297.

[31] Fafchamps, M., and F. Gubert (2007) The Formation of Risk Sharing Networks, *Journal of Development Economics*, 83: 326-350.

[32] Geweke, J., and M. Keane (2001) Computationally Intensive Methods for Integration in Econometrics, in J. J. Heckman and E. Leamer eds., *Handbook of Econometrics*, Volume 5, Chapter 56: 3463-3568.

[33] Goyal, S. (2007) *Connections: An Introduction to the Economics of Networks*, Princeton University Press, Princeton, NJ.

[34] Goyal, S., and S. Joshi (2006) Unequal Connections, *International Journal of Game Theory*, 34: 319-349.

[35] Goyal, S., and F. Vega-Redondo (2007) Structural Holes in Social Networks, *Journal of Economic Theory*, 137: 460-492.

[36] Graham, B. (2016) An Econometric Model of Link Formation with Degree Heterogeneity, *Econometrica* (Conditional acceptance).

[37] Hajivassiliou, V. A., and P. A. Ruud (1994) Classical Estimation Methods for LDV Models Using Simulation, in R. F. Engle and D. L. McFadden eds., *Handbook of Econometrics*, Volume 4, Chapter 40: 2384-2443.

[38] Hellmann, T. (2012) On the Existence and Uniqueness of Pairwise Stable Networks, *International Journal of Game Theory*, forthcoming.

[39] Hirano, K., and J. R. Porter (2003) Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support, *Econometrica*, 71(5), 1307–1338.

[40] Jackson, M. O. (2008) *Social and Economic Networks*, Princeton University Press, Princeton, NJ.

[41] Jackson, M. O., T. Barraquer, and X. Tan (2012) Social Capital and Social Quilts: Network Patterns of Favor Exchange, *American Economic Review*, 102 (5): 1857-1897.

[42] Jackson, M. O., and A. van der Nouweland (2005) Strongly stable networks, *Games and Economic Behavior*, 51: 420-444.

[43] Jackson, M. O., and B. W. Rogers (2007) Meeting Strangers and Friends of Friends: How Random Are Social Networks? *American Economic Review*, 97(3): 890-915.

[44] Jackson, M. O., and A. Watts (2001) The Existence of Pairwise Stable Networks, *Seoul Journal of Economics*, 14, 3: 299-321.

[45] Jackson, M. O., and A. Watts (2002) The Evolution of Social and Economic Networks, *Journal of Economic Theory,* 106: 265-295.

[46] Jackson, M. O., and A. Wolinsky (1996) A Strategic Model of Social and Economic Networks, *Journal of Economic Theory*, 71: 44-74.

[47] Kallenberg, O. (2005) *Probabilistic Symmetries and Invariance Principles*, Springer, New York, NY.

[48] Kline, B. and E. Tamer (2015) Bayesian Inference in a Class of Partially Identified Models, working paper.

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

[50] Lovász, L. and B. Szegedy (2006) Limits of Dense Graph Sequences, *Journal of Combinatorial Theory*, Series B 96: 933-957.

[51] Lovász, L. (2012). *Large Networks and Graph Limit*, American Mathematical Society Colloquium Publications, Volumn 60.

[52] Mayer A., and S. L. Puller (2008) The Old Boy (and Girl) Network: Social Network Formation on University Campuses, *Journal of Public Economics*, 92: 329–347.

[53] McFadden, D. (1989) A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, *Econometrica*, 57(5): 995-1026.

[54] Mele, A. (2011) A Structural Model of Segregation in Social Networks, working paper.

[55] Menzel, K. (2015) Large Matching Markets as Two-Sided Demand Systems, *Econometrica*, 83(3): 897-941.

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

[57] Menzel, K. (2016b) Strategic Network Formation with Many Agents, working paper.

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

[59] Milgrom, P. and J. Roberts (1990) Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities, *Econometrica*, 58(6): 1255-1277.

[60] Moretti, E. (2011) Social Learning and Peer Effects in Consumption: Evidence from Movie Sales, *Review of Economic Studies*, 78, 356-393.

[61] Monderer, D. and L. S. Shapley (1994) Potential Games, *Games and Economic Behavior*, 14: 124-143.

[62] Myerson, R. (1991) *Game Theory: Analysis of Conflict*. Harvard University Press.

[63] Nakajima, R. (2007) Measuring Peer Effects on Youth Smoking Behavior, *Review of Economic Studies*, 74(3): 897-935.

[64] Neal, R. (2003) Slice Sampling, *Annals of Statistics*, 31(3): 705-767.

[65] Orlin J., A. Punnen, and A. Schulz (2004) Approximate Local Search in Combinatorial Optimization, *SIAM Journal on Computing*, 33(5): 1201-1214.

[66] Pakes, A., and D. Pollard (1989) Simulation and the Asymptotics of Optimization Estimators, *Econometrica,* 57(5): 1027-1057.

[67] Pakes, A., J. Porter, K. Ho, and J. Ishii (2006) Moment Inequalities and Their Application, working paper.

[68] Ridder, G., and S. Sheng (2016) Estimation of Large Network Formation Games, working paper.

[69] Romano, J. P., and A. M. Shaikh (2010) Inference for the Identified Set in Partially Identified Econometric Models, *Econometrica*, 78(1): 169-211.

[70] Snijders, T. (2002) Markov Chain Monte Carlo Estimation of Exponential Random Graph Models, *Journal of Social Structure*, 3(2).

[71] Tamer, E. (2003) Incomplete Simultaneous Discrete Response Model with Multiple Equilibria, *Review of Economic Studies*, 70: 147-190.

[72] Topkis, D. M. (1979) Equilibrium Points in Nonzero-Sum n-Person Submodular Games, *SIAM Journal of Control and Optimization*, 17(6): 773-787.

[73] van der Vaart, A., and J. A. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.

[74] Young, H. P. (1993) The Evolution of Conventions, *Econometrica*, 61(1): 57-84.