A Random-Field Approach to Inference in Large Models of Network Formation

Michael P. Leung*

July 11, 2015

Note: Superceded by "A Weak Law for Moments of Pairwise-Stable Networks."

Abstract. We establish a law of large numbers and central limit theorem for a large class of network statistics, enabling inference in models of network formation with only a single network observation. Our model allows the decision of an individual (node) to form associations (links) to depend quite generally on the endogenous structure of the network. Formally, we prove that certain node-level functions of the network constitute α-mixing random fields, objects for which central limit theorems exist. The key assumptions are that (1) nodes endowed with similar attributes prefer to link (homophily); (2) there is enough "diversity" in node attributes; and (3) certain latent sets of nodes that are unconnected form their links independently ("isolated societies" do not "coordinate"). Our results enable the estimation of certain network moments that are useful for inference. We leverage these moments to construct moment inequalities that define bounds on the identified set. Relative to feasible alternatives, these bounds are sharper and computationally tractable under weaker restrictions on network externalities.

JEL Codes: C31, C57, D85, L14
Keywords: social networks, network formation, multiple equilibria, large-market asymptotics, random fields, α-mixing

*Department of Economics, Stanford University. E-mail: mpleung@stanford.edu. I thank my advisors, Arun Chandrasekhar, Matt Jackson, and especially Han Hong, for their advice, criticism, and support and many helpful discussions. I also thank Marinho Bertanha, Tim Bresnahan, Michael Dickstein, Jessie Li, Xing Li, Josh Mollner, Frank Wolak, and participants at the Stanford structural I.O. lunch for useful feedback. I gratefully acknowledge support from a B. F. Haley and E. S. Shaw Fellowship through the Stanford Institute for Economic Policy Research and a Ric Weiland Graduate Fellowship.

1 Introduction

We develop asymptotic theory for inference in network-formation models, which have important applications for the study of social networks. The economic perspective on network science is broadly concerned with two distinct but related questions: (1) what incentives govern the formation of social connections, and (2) how do these connections influence economic behavior? To empirically study (1), we require a method to estimate models of network formation. For example, a well-known determinant of network formation is homophily, the notion that two individuals are more likely to associate if their attributes are similar. Another possible determinant is popularity, meaning that individuals prefer to be friends with those who already have many friends. Here, the endogenous linking decisions of others, rather than their attributes, directly influence an individual's linking decision. Network-formation models enable the econometrician to disentangle endogenous determinants of link formation, such as popularity, from exogenous determinants, such as homophily.

The peer-effects literature is concerned with question (2), how networks mediate social interactions. A pervasive problem is network endogeneity or *latent homophily* (Shalizi and Thomas, 2011): links tend to form between individuals with similar *unobserved* attributes. For example, high-ability students are likely to become friends, which confounds the effect of high-achieving students on peer achievement. Network-formation models can provide a solution for latent homophily, as formally modeling the incentives for forming social connections can control for network endogeneity (Badev, 2013; Goldsmith-Pinkham and Imbens, 2013; Uetake, 2012).

This paper studies the following empirical model of network formation:

$$G_{ij} = 1 \quad \text{if and only if} \quad V_{ij}^n(G, W; \theta_0) \geqslant 0, \tag{1}$$

for all $i, j \in \mathcal{N}_n$, the set of n nodes/individuals/agents. Here G is the observed network, represented as an $n \times n$ matrix whose ijth entry, G_{ij}, is an indicator for whether or not i and j form a link. The set of node attributes is represented by W. This model is a natural extension of the standard discrete-choice model that allows $V_{ij}^n(\cdot)$ to depend on the endogenous outcome G, what we refer to as network externalities. For example, if a link represents a friendship, two individuals may be more likely to become friends if they both have a friend in common, in which case link formation depends on the presence of other links. Model (1) has been used in empirical network studies (e.g. Comola, 2012; DeWeert, 2004; Powell et al., 2005), but the formal asymptotic properties of their inference procedures have not been studied. More commonly, empirical researchers fit dyadic regression models (e.g. Bramoullé and Fortin, 2010; Fafchamps and Gubert, 2007), which are special cases of (1) in which $V_{ij}^n(\cdot)$ does not depend on G. This reduces the model to a standard discrete-choice setup but at the cost of ruling out externalities *a priori*. The goal of this paper is to develop the formal statistical theory for discrete-choice models of type (1) to enable inference on network externalities.

The existence of network externalities creates two fundamental difficulties for inference. First, researchers typically observe a small number of networks in the data, either because it is costly to acquire data on many networks or because the network being studied only occurs in nature as a single large graph. In this case, treating the network as the unit of observation is inappropriate, since this corresponds to a sample size of one. Instead, we would like to equate a large network with a large sample size by treating the *node* as the unit of observation. This formally corresponds to asymptotics under a sequence of models that sends the size of the network to infinity ("large-market" asymptotics), rather than the number of networks ("many-market" asymptotics). This is a challenging statistical problem because network externalities induce correlation between links, and limit theory usually requires some form of independence for a central limit theorem to be valid.

The second fundamental difficulty is that the presence of network externalities typically renders the model incomplete because for any draw of node attributes W, there may be multiple networks G that satisfy the system of equations implied by (1). This is the same problem we face in econometric supply and demand models in which supply is linear and demand is backward-bending. Here, the system of simultaneous equations predicts two possible market equilibria, and the equations alone are uninformative for the realized equilibrium. Thus, in order to complete the model, we must specify a selection mechanism that maps this set of possible outcomes to a unique outcome, but theory provides little prior information on the form of the selection mechanism. Without a formal model of equilibrium selection, the inference procedure must account for the selection mechanism as an unknown nuisance parameter.

Our solution to the first difficulty is to propose conditions under which network dependence is sufficiently limited, establishing a law of large numbers and central limit theorem, which are basic building blocks for inference. Define node i's *node statistic* to be a function $\psi_i(G)$ that depends on G only through the subnetwork on the *component* of node i, that is, the set of nodes that are *connected* (directly or indirectly) to node i.[1] Many sample analogs of network moments that are useful for inference can be written as averages $\frac{1}{n}\sum_{i=1}^n \psi_i^n(G)$, but showing the consistency of these estimators for their respective moments is nontrivial. A simple example is the average degree, the average number of direct connections a node forms:

$$\frac{1}{n}\sum_{i=1}^n \underbrace{\sum_{j\neq i} G_{ij}}_{\psi_i^n(G)}\,.$$

Assuming expected degree is uniformly bounded, it is natural to expect that a law of large numbers holds, and this average converges in probability to its expectation, as $n \to \infty$. However, this is only obvious if links are independent, which is not necessarily true if network externalities exist. We derive conditions under which a law of large numbers is valid. These conditions ensure that $\{\psi_i(G); i \in \mathcal{N}_n\}$ constitutes an α-mixing random field, which formally means that the linking decisions of different nodes are sufficiently uncorrelated.

A random field is a collection of random variables indexed by $\mathbb{R}^d$, with the special case of time series corresponding to $d = 1$. Informally, a time series is *mixing* if observations far apart in time are less correlated. Analogously, a random field is mixing observations far apart in *space* are less correlated. If nodes are positioned in some space, say, corresponding to geographic location, or any other set of attributes, then three main conditions are needed for mixing. First, nodes are homophilous in distance, meaning that nodes that are far apart in space (i.e. different in terms of attributes) are unlikely to link. Second, the network is "diverse," meaning that enough nodes are sufficiently far apart in space. Third, for a given W and θ_0, latent sets of nodes that are not connected under any equilibrium network must form their links independently, i.e. "isolated societies" of nodes do not "coordinate" on their linking decisions. We formalize these conditions in the main text.

[1] Two nodes i and j are *connected* if there exists a path from i to j. A *path* from i to j is a distinct sequence of nodes starting with i and ending with j such that for each k, k' in this sequence, $G_{k,k'} = 1$. For directed networks, this definition of connectivity corresponds to i and j being *weakly connected*.

The second fundamental difficulty results from the fact that for any given W and θ_0, multiple networks may satisfy the system of equations (1). Without further restrictions, this typically implies that the structural parameter θ_0 is only partially identified, meaning that there may be a set of parameters consistent with the model and the data, rather than a singleton. We derive moment inequalities implied by the model that define an identified set. These moments are computationally tractable and can be used to estimate the identified set and construct confidence intervals for θ_0. In particular, computation of the moments avoids a common curse of dimensionality in the number of nodes (discussed in related literature below) under weaker restrictions on network externalities than feasible alternatives in the literature. The cost of computational tractability is that the identified set is conservative. However, it is shown to be sharper than some existing alternatives.

Related literature. The literature on estimating models of network formation is largely in its infancy. Central themes include (1) inference with only a single network observation, (2) models with multiple equilibria, and (3) computational challenges. The latter largely stem from the fact that the number of possible networks on n nodes is on the order of 2^{n^2}, which exceeds the number of elementary particles in the universe for $n > 30$.

There is a growing econometric literature on estimating models of network formation related to (1). The important work of Boucher and Mourifié (2013) appears to be the first to suggest leveraging the machinery of random fields for estimating network-formation models using a single network observation. However, they assume the model has a unique equilibrium, which strongly restricts the class of estimable models. For instance, none of the examples considered in this paper satisfy this assumption. Our results also substantially broaden the class of estimable network statistics beyond the score function, which is their primary focus. Other papers that provide frequentist estimation procedures for related models when only a single network is observed are Dzemski (2014), Graham (2014), and Leung (2014). Both Dzemski and Graham consider models without externalities but allow for unrestricted unobserved heterogeneity. Leung considers models with incomplete information.

Miyauchi (2013) and Sheng (2014) provide moment inequalities for inference in model (1). Both assume that the econometrician observes many independent networks. Miyauchi's moments require computing a set of equilibrium networks, which suffers from a curse of dimensionality in the number of nodes and therefore is only feasible for small networks. Sheng is the first to derive moment inequalities that avoid this curse of dimensionality under a novel locality restriction on network externalities. We show that there exists a much larger set of computable moment inequalities that includes those developed by Sheng and therefore collectively defines a sharper identified set. We also show that the local externalities restriction can be substantially weakened. While Sheng assumes that the latent index $V_{ij}^n(\cdot)$ depends on nodes at most $\ell = 2$ links away from i or j, we provide a condition under which any $\ell < \infty$ is feasible, significantly expanding the class of estimable models. Additionally, our estimation procedure has the advantage of not requiring a nonparametric density estimator.

A large statistics literature studies inference for random graph models, which are alternative approaches to modeling network formation. In this literature, exponential random graph models (ERGMs) are the leading models for studying network externalities. These models are typically estimated using Markov Chain Monte Carlo (MCMC), but this procedure suffers from a curse of dimensionality in the number of nodes and is therefore computationally infeasible in nontrivial models for networks even of moderate size (Bhamidi et al., 2011). Formal asymptotic theory for estimating ERGMs has recently been developed by Chandrasekhar and Jackson (2014), who consider the case of a single network observation. Importantly, they derive a new inference procedure that avoids the curse of dimensionality faced by MCMC.

There is also an econometric literature on estimating dynamic network-formation models, which augment model (1) with a meeting technology that determines how nodes form links sequentially over time. In some sense, this can be thought of as a restriction on the equilibrium selection mechanism. Much of this literature focuses on inference when the econometrician observes a single network, and most adopt Bayesian estimation techniques for this purpose (Christakis et al., 2010; Hsieh and Lee, 2012; Mele, 2013). The models studied in the latter two papers induce likelihoods that reduce to ERGMs, and consequently, their inference procedures based on MCMC are limited by a curse of dimensionality. Chandrasekhar and Jackson (2014) show that a class of dynamic models of network formation can microfound the random graph models they study. Relative to our approach, these models require stronger separability assumptions on the latent index $V_{ij}^n(\cdot)$ and do not allow for unobserved heterogeneity. On the other hand, they do not require many of the conditions we impose for valid inference, such as homophily or diversity.

Our model is related to the large literature on estimating games of complete information (e.g. Bajari et al., 2010; Bresnahan and Reiss, 1990; Ciliberto and Tamer, 2009; Galichon and Henry, 2011; Tamer, 2003). Several of these papers develop inference procedures based on moment inequalities (Beresteanu et al., 2011; Galichon and Henry, 2011; Henry et al., forthcoming), but most require computing the set of equilibria, which is computationally infeasible for even moderately large games. Our procedure avoids having to compute this set. Most papers in this literature consider many-market asymptotics, but there are several exceptions (Agarwal and Diamond, 2014; Bisin et al., 2011; Fox, 2010; Menzel, 2014a,b; Song, 2012). These papers either study matching or peer-effects models, which do not include models of network formation with network externalities.

In the next section, we discuss our model and an overview of our approach in the context of two motivating examples. We discuss in detail the economic interpretation of the assumptions we impose on the model to limit the extent of network dependence. Section 2 formally presents these assumptions and the main result, a law of large numbers and central limit theorem for network statistics. We develop our inference procedure in section 3, first discussing the construction of moment inequalities and then turning to implementation. Section 4 concludes.

Notation. For any set A, $|A|$ denotes its cardinality, 2^A the set of all subsets of A, and A^c the complement of A. For any sets S_0 and S_1 with $S_0 \subseteq S_1$, $S_1 \backslash S_0 = \{s \in S_1 : s \notin S_0\}$. For $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $||x||$ is the supremum norm of x, equal to $\max_{i \in \{1,\dots,d\}} |x_i|$. Random variables are represented using upper-case letters and their realizations using lower-case letters.

2 Asymptotics for Network Statistics

Let $\mathcal{N}_n = \{1, \dots, n\}$ be the set of node labels. A node i is endowed with a position $\rho_i \in \mathbb{R}^{d_\rho}$ and attributes $X_i \in \mathbb{R}^{d_x}$. Positions are special attributes in which we will require nodes to be homophilous, such as geographic location. For this purpose, let $\delta_{ij} = (|\rho_{i1} - \rho_{j1}|, \dots, |\rho_{id_\rho} - \rho_{jd_\rho}|)$. We interpret $||\delta_{ij}||$ as the *dissimilarity* between nodes with respect to positions, which will play a crucial role in obtaining a central limit theorem. Positions are considered *pre-determined*, and assumptions in Section

2 will ensure that positions are sufficiently scattered in $\mathbb{R}^{d_\rho}$.[2]

Each pair of nodes is endowed with a vector of pair-specific attributes $\zeta_{ij} \in \mathbb{R}^{d_\zeta}$, e.g. an idiosyncratic pair-level shock that captures meeting opportunities. We define $W = \big((\rho_i, \rho_j, X_i, X_j, \zeta_{ij}); i, j \in \mathcal{N}_n\big)$, the set of all node primitives. For $A \subseteq \mathcal{N}_n$, let $W_A = \big((\rho_i, \rho_j, X_i, X_j, \zeta_{ij}); i, j \in A\big)$, the submatrix of W associated with nodes in A. Let $Z_{ij} = (X_i, X_j, \zeta_{ij})$, the vector of "exogenous factors" associated with pair (i, j). We require a standard restriction on the joint distribution of attributes.

Assumption 1. *For any n and $A, A' \subseteq \mathcal{N}_n$ disjoint, $W_A \perp\!\!\!\perp W_{A'}$.*

The econometrician observes positions and subvectors of X_i and ζ_{ij}, denoted X_i^o and ζ_{ij}^o. We denote the unobserved remainders of X_i and ζ_{ij}, respectively, by X_i^u and ζ_{ij}^u. The econometrician also observes a network $G \in \mathbf{G}_n$, the set of $n \times n$ matrices with 0-1 entries. We call G_{ij}, the ijth entry of G, the *potential link* between i and j.

We study the latent-index model of link formation

$$G_{ij} = 1 \quad \text{if and only if} \quad V_{ij}^n(G, W; \theta_0) \geqslant 0,$$

where $V_{ij}^n(\cdot)$ is known up to the parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_t}$.[3] In this model, a link forms between i and j if and only if the pair's latent index $V_{ij}^n(\cdot)$ is nonnegative. In the case of undirected networks, this index can be interpreted as the joint surplus that "agents" i and j jointly enjoy from the addition of link G_{ij}, taking as given the rest of the network. Hence, (1) is a generalization of discrete-choice models that allows the formation of a link between i and j to depend on the presence of links between other nodes. We will refer to any network G that satisfies the inequalities defined by (1) for a given W and θ as an *equilibrium.*

Clearly, the model can be microfounded as a game of complete information in which the solution concept is *pairwise-stable equilibrium with transfers*; see e.g. Sheng (2014), Definition 2. The analysis in this paper can easily be extended to non-transferable utility, as well. In the case of directed networks, (1) is a consequence of Nash equilibrium play in a static link announcement game (Myerson, 1977).

[2] Non-random positions is the standard framework used in the literature on random fields. This is analogous to time-series settings where random variables are associated with fixed positions in time.

[3] When the network is undirected, it is sensible to assume $V_{ij}^n(G, W; \theta_0) = V_{ji}^n(G, W; \theta_0)$ and $\zeta_{ij} = \zeta_{ji}$.

Our main object of interest are averages of certain node statistics, which serve as estimators for subnetwork moments used for inference. In order to define these statistics, we first need some additional notation. The *component* of a network G is subnetwork such that any two nodes are connected in G. For a given G, denote by $C_i(G)$ the size of the maximal component that contains i, which is the largest set of nodes containing i that forms a network component. For brevity, we often refer to $C_i(G)$ simply as i's component. Also let $C_{ij}(G) = C_i(G) \cup C_j(G)$. For any $A \subseteq \mathcal{N}_n$, define G_A as the submatrix of G corresponding only to links between nodes in A and $\mathbf{G}_A$ as the set of labeled networks on A. Let $G_{-A} \equiv \{G_{ij} : i,j \in \mathcal{N}_n, \text{ not both in } A\}$ be the set of links in G, excluding links between nodes in A. For $i,j \in \mathcal{N}_n$, define $\{i \leftrightarrow j\}$ as the event that i and j are connected in network G, leaving the network implicit in the definition.

With these definitions in hand, we can associate with each $i \in \mathcal{N}_n$ a *node statistic* $\psi_i^n \equiv \psi_i^n(G,W)$ for some known function ψ_i^n such that $\psi_i^n(G,W) = \psi_i^n(G',W')$ for any $(G,W),(G',W')$ such that $G_{C_i(G)} = G'_{C_i(G')}$ and $W_{C_i(G)} = W'_{C_i(G')}$. In other words, ψ_i^n only depends on the links and the attributes of the nodes in i's component. We aim to provide conditions under which the triangular array $\{\psi_i^n; i \in \mathcal{N}_n, n \in \mathbb{N}\}$ is an α-mixing random field. This ensures that averages $\frac{1}{n}\sum_{i=1}^n \psi_i^n$, which we term *network statistics*, are consistent for their expectations and asymptotically normal by a law of large numbers and central limit theorem for random fields, results that are necessary for inference on θ_0.

In a directed network in which links represent, say, lending relationships, some simple examples of moments we may like to estimate include $\mathbf{P}(G_{ij}=1, G_{ji}=0)$ (unidirectional lending) and $\mathbf{P}(G_{ij}=G_{ji}=1)$ (bidirectional lending), which intuitively help to identify *reciprocity*, the intrinsic tendency for i to lend to j if j lends to i. Their sample analogs are examples of network statistics $\frac{1}{n}\sum_i \psi_i(G)$ mentioned in the introduction, where $\psi_i(G)$ only depends on G through i's network component. For example, in the case of the bidirectional referral probability, the node statistic $\psi_i(G) = \frac{1}{2}\sum_j G_{ij}G_{ji}$, and $\mathbf{E}[\frac{1}{2n}\sum_{i,j} G_{ij}G_{ji}] = \frac{1}{n}\sum_{i,j>i} \mathbf{P}(G_{ij}=G_{ji}=1)$.[4]

With many independent networks, these moments are easily estimated by their sample analogs. However, with only a single network, consistency requires sufficient

[4]Note that even though $\psi_i(G)$ is a sum over n elements, it does not grow arbitrarily large as $n \to \infty$ because the network is sparse, both empirically in most social networks and theoretically in a sense described later.

independence in the network for a law of large numbers. This holds if the node statistics $\psi_i(G, W)$ are uncorrelated enough in the sense of α-mixing, which we now define.

Let $\{\psi_i^n; i \in \mathcal{N}_n, n \in \mathbb{N}\}$ be a triangular array on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. For $A, A' \subseteq \mathcal{N}_n$, let $d(A, A') = \inf\{||\delta_{ij}||; i \in A, j \in A'\}$, and let $\sigma(\psi_i^n; i \in A)$ denote the σ-algebra generated by the vector $(\psi_i^n; i \in A)$. Define

$$\alpha_n(A, A') = \sup\{|\mathbf{P}(H \cap H') - \mathbf{P}(H)\mathbf{P}(H')|; H \in \sigma(\psi_i^n; i \in A), H' \in \sigma(\psi_i^n; i \in A')\},$$
$$\alpha_{a,a',n}(r) = \sup\left\{\alpha_n(A, A'); |A| \leqslant a,\ |A'| \leqslant a',\ d(A, A') \geqslant r\right\},$$
$$\bar{\alpha}_{a,a'}(r) = \sup_n \alpha_{a,a',n}(r).$$

The mixing coefficient $\alpha_n(A, A')$ is a measure of dependence between the node statistics of nodes in A and A'. This measure lies in $[0, 1]$, with zero corresponding to independence and departures from zero signifying dependence. For a CLT, we need $\bar{\alpha}_{a,a'}(r) \to 0$ at a fast enough rate as the dissimilarity r goes to infinity. This is the spatial analog of the usual idea in time series that observations that are distant in time should be close to independent.

Definition 1 (α-Mixing)**.** The triangular array $\{\psi_i^n; i \in \mathcal{N}_n, n \in \mathbb{N}\}$ is an *α-mixing random field* if its α-mixing coefficient satisfies $\bar{\alpha}_{a,a'}(r) \leqslant C(a + a')^{\chi}\hat{\alpha}(r)$ with $\hat{\alpha}(r) = O(r^{-d_\rho(2\eta+1)-\kappa})$ for some $C < \infty$, $\eta \geqslant \max\{\chi, 1\}$, $\kappa > 0$.

The central difficulty is that with network externalities, links can be highly dependent, meaning that the mixing coefficient is far from zero for most pairs. We next detail a set of conditions that are sufficient for the set of node statistics to constitute an α-mixing random field.

2.1 Latent Index

The first set of assumptions are restrictions on the latent index $V_{ij}^n(\cdot)$.

Assumption 2 (Specification)**.** *There exists $E \subseteq \mathbb{R}^{d_\mathcal{E}}$ such that the following hold.*

(a) For any $n \in \mathbb{N}$ and $i, j \in \mathcal{N}_n$, there exists $\mathcal{E}_{ij}^n(\cdot,\cdot)$ with range E such that $V_{ij}^n(G, W; \theta) = V_{ij}(\delta_{ij}, \mathcal{E}_{ij}^n(G, W), Z_{ij}; \theta)$.

(b) *Let E_k be the kth dimension of E. If for all $i, j \in \mathbb{N}$, $V_{ij}(\cdot)$ is increasing in the kth component of $\mathcal{E}^n_{ij}(G, W)$, then E_k is bounded above. Otherwise, E_k is bounded.*

For ease of notation, we will often abbreviate $\mathcal{E}^n_{ij} = \mathcal{E}^n_{ij}(G, W)$.

Assumption 2 requires that the latent index for nodes i and j is a function of a finite-dimensional vector of factors $(\delta_{ij}, \mathcal{E}^n_{ij}, Z_{ij})$. We refer to $\mathcal{E}^n_{ij}$ as *endogenous factors* and to Z_{ij} as *exogenous factors*, so called because the former depend on the endogenous network G, while the latter do not. The most substantive requirement is Assumption 2(b), that the endogenous factors are uniformly bounded, which is needed to ensure that the latent index does not diverge to infinity with n. If $V_{ij}(\cdot)$ is increasing in an endogenous factor for all pairs (i, j), then we only require boundedness above. This allows for cost functions that depend on G. For example, it is often reasonable to impose a *capacity constraint*, which is the constraint that no node has degree exceeding $\bar{D} < \infty$. This can be modeled, for example, as a function

$$c(G_i) = \begin{cases} 0 & \text{if } \sum_j G_{ij} \leqslant \bar{D} \\ -\infty & \text{if otherwise} \end{cases} \tag{2}$$

where G_i is the vector of links $(G_{ij}; j \in \mathcal{N}_n)$. If $c(G_i)$ is a component of $\mathcal{E}^n_{ij}$, then it is natural to assume V_{ij} is increasing in $c(\cdot)$. Evidently, $c(\cdot)$ is bounded above, so this satisfies Assumption 2(b).

Assumption 2 is satisfied by many endogenous factors of interest, such as $\mathbf{1}\{i \leftrightarrow j\}$. Other factors, such as the number of common friends that i and j share, $\sum_k G_{ik}G_{jk}$, may be bounded above under a capacity constraint. If such a constraint is inappropriate, then an alternative is to scale factors by the rate at which they diverge. For instance, while i's degree violates the assumption when degree diverges at rate n, i's average degree, $\frac{1}{n}\sum_j G_{ij}$, does not. Even with a capacity constraint, uniform boundedness does rule out endogenous factors such as the total income of i's friends, $\sum_{j=1}^n G_{ij}M_j$, where M_j is the income of node j, if one is unwilling to assume that income has bounded support. We next provide additional examples of endogenous factors.

Example 1. Suppose G represents a friendship network, and $G_{ij} = 1$ if individuals i and j are friends. Let M_i be individual i's income, A_i her age, and α_i her latent gregariousness. The econometrician only observes the subvector (M_i, A_i) for each i.

Define $c^*(G_i, G_j) = c(G_i) + c(G_j)$, where $c(\cdot)$ is defined in (2). The model is

$$V_{ij}^n(G, W; \theta) = \theta_1 + \theta_2|M_i - M_j| + \theta_3(A_i + A_j) + \theta_4\mathbf{1}\{\exists\, k : G_{ik} = G_{jk} = 1\} \\ + \theta_5 \sum_{k \neq i,j} G_{kj} + c^*(G_i, G_j) + \alpha_i + \alpha_j + \zeta_{ij}.$$

The parameter θ_2 captures homophily in income. The parameter θ_4 captures *transitivity* or *clustering*, the tendency for individuals with friends in common to become friends. The importance of transitivity is widely recognized (see e.g. Christakis et al., 2010; Goldsmith-Pinkham and Imbens, 2013; Jackson, 2008). The parameter θ_5 represents the importance of popularity; if $\theta_5 > 0$, then individuals prefer to be friends with those who have many friends. The cost function enforces a capacity constraint, limiting the total number of links a node can form. This is intuitive, since we expect that people have a finite capacity to form friendships. Finally, the random effects α_i and α_j allow for *degree heterogeneity* (Graham, 2013), the unobserved tendency for some individuals to form more links than others.

The next assumption says that $\mathcal{E}_{ij}^n$ only depends on G and W through the subnetwork formed on the components of i and j and the attributes of these nodes.

Assumption 3 (Component Externalities)**.** *Let G, W and G', W' satisfy $G_{C_{ij}(G)} = G'_{C_{ij}(G')}$ and $W_{C_{ij}(G)} = W'_{C_{ij}(G')}$. Then $\mathcal{E}_{ij}^n(G, W) = \mathcal{E}_{ij}^n(G', W')$ for any $i, j \in \mathcal{N}_n$ and $n \in \mathbb{N}$.*

This is a commonly satisfied requirement. The model in Example 1 satisfies this condition, since $V_{ij}(\cdot)$ only depends on nodes at most one link away, whereas component externalities allows $V_{ij}(\cdot)$ to depend on nodes any finite number of links away. This condition also permits the latent index to depend on the path distance between two nodes or a node's Bonacich centrality. Generally, Assumptions 2 and 3 are weak and allow for a broad spectrum of network externalities, including those studied in Boucher and Mourifié (2013), Christakis et al. (2010), Goldsmith-Pinkham and Imbens (2013), Mele (2013), Miyauchi (2013), and Sheng (2014).

The next assumption serves two purposes. First, it imposes a restriction on the tails of the distributions of attributes, essentially requiring that large realizations of attributes are rare. Second, it formally defines homophily.

Assumption 4. *There exist $\tau : \mathbb{R}_+ \to \mathbb{R}_+$, $\mathcal{C}_x : \mathbb{R}_+ \to 2^{\mathbb{R}^{d_x}}$, and $\mathcal{C}_\zeta : \mathbb{R}_+ \to 2^{\mathbb{R}^{d_\zeta}}$ such that $\mathcal{C}_x$ and $\mathcal{C}_\zeta$ are nonempty, nondecreasing, set-valued functions and the following conditions hold.*[5]

(a) (Tails) For all r,

$$\max\left\{\sup_n \max_{i\in\mathcal{N}_n} \mathbf{P}\big(X_i \notin \mathcal{C}_x(r)\big),\ \sup_n \max_{i,j\in\mathcal{N}_n} \mathbf{P}\big(\zeta_{ij} \notin \mathcal{C}_\zeta(r)\big)\right\} \leqslant \tau(r),$$

and $\tau(r)r^{5d_\rho+\varphi} \overset{r\to\infty}{\longrightarrow} c \geqslant 0$ for some $\phi > 0$.

(b) (Homophily) Define $\mathcal{C}(\cdot) = \mathcal{C}_x(\cdot) \times \mathcal{C}_x(\cdot) \times \mathcal{C}_\zeta(\cdot)$. For any $\mathbb{R}^{d_\rho}$-valued sequence $\{\delta_m\}_{m=0}^\infty$ with $||\delta_m|| \to \infty$ and any $\theta \in \Theta$,

$$\limsup_{m\to\infty} \max_{i,j\in\mathbb{N}} \sup_{\mathcal{E}\in E} \sup_{z\in\mathcal{C}(||\delta_m||)} V_{ij}\big(\delta_m, \mathcal{E}, z; \theta\big) < 0,$$

The function τ controls the rarity of large draws of ζ_{ij} and X_i, where largeness is measured by the complement of the set $\mathcal{C}(\cdot)$. The requirement that $\tau(r)r^{5d_\rho+\varphi} \to c \geqslant 0$ formalizes rarity; the tails of ζ_{ij} cannot be too heavy in the sense of containing too much probability mass. For example, if ζ_{ij} is scalar and its distribution belongs in the exponential family, then part (a) is satisfied for $\mathcal{C}_\zeta(r) = [-r, r]$ and $\tau(r) = Ce^{-r}$ for some $C > 0$. Hence, if ζ_{ij} captures idiosyncratic meetings between individuals that predispose them to form friendships, then this means that such meetings occur infrequently in the population, which is sensible given that in most social networks, the typical number of friends is much smaller than the universe of potential friends. This condition is important because large draws of ζ_{ij} work against homophily; pairs with a high pair-specific value from linking may form a connection even if they are extremely dissimilar. Since our goal is to limit network dependence by restricting the connectivity of the network through homophily, it is therefore important to control the probability of large realizations of ζ_{ij}.

Assumption 4(b) captures homophily, the principle that "similarity breeds connection," which is ubiquitous in social networks (McPherson, Smith-Lovin and Cook, 2001). It is formalized here by requiring that the latent index diverges to negative infinity with dissimilarity. Hence, nodes are disinclined to link with highly dissimilar nodes. Note that in Example 1, a necessary condition for homophily is $\theta_2 < 0$.

[5] A set-valued function $\mathcal{C} : \mathbb{R} \to 2^{\mathbb{R}^k}$ is nondecreasing if $r < r'$ implies $\mathcal{C}(r) \subseteq \mathcal{C}(r')$.

Homophily relates to the tails condition due to the presence of $\mathcal{C}(\cdot)$ in the supremum. First consider the simple case in which X_i and ζ_{ij} have bounded support. Then we can take $\mathcal{C}(\cdot)$ to be the Cartesian product of their supports, so that $\mathcal{C}(r)$ is identical for all r. Part (b) says that highly dissimilar nodes prefer not to link even if the endogenous and exogenous factors take on values in their supports that are most favorable for link formation. Thus, as two nodes move apart, there must a point at which the "disutility" from being too far apart dominates all other incentives for link formation. The purpose of the set-valued function $\mathcal{C}(\cdot)$ is to allow Z_{ij} to have components with full support. In part (b), we restrict the values of Z_{ij} to be at their most favorable for link formation within the constraint set $\mathcal{C}(\cdot)$. This set can grow with dissimilarity δ_{ij} but not so quickly that it overtakes the effect of homophily. Part (a) ensures that Z_{ij} lies in this set with high probability. Hence, the idea is to ensure that large realizations of the attributes are rare in order for homophily to dominate.

2.2 Diversity

The next pair of assumptions require nodes to be sufficiently dissimilar from one another in terms of positions. For $n \in \mathbb{N}$ and $i \in \mathcal{N}_n$, let $\mathcal{B}_i(r) = \{j : j \in \mathcal{N}_n, ||\delta_{ij}|| < r\}$.

Assumption 5 (Increasing Domain)**.** *For some $r_0 > 0$, $\sup_n \max_{i \in \mathcal{N}_n} \left|\mathcal{B}_i(r_0)\right| < \infty$. Without loss of generality, normalize $r_0 = 1$.*

This is a generalization of the increasing-domain condition used in asymptotics for random fields (e.g. Assumption 1 of Jenish and Prucha, 2009) that allows a small number nodes to have the same position but requires most pairs of nodes to have a minimum amount of dissimilarity r_0. Increasing-domain asymptotics in turn generalizes the usual assumption in time series that time periods are located in $\mathbb{N}$, which ensures that observations are spread out in time, rather than packed in a fixed interval. In part, this assumption ensures that homophily (Assumption 4(b)) is meaningful, since it would play little role in link formation if all nodes had similar positions.

Remark 1 (Network Sparsity)**.** We show in Proposition B.1 that under Assumptions 1, 2, 4, and 5, the average expected degree is uniformly bounded over n. That is, of all the potential links a node may form, the fraction of links that do form is vanishing in n. Intuitively, increasing domain ensures that nodes are spread out in $\mathbb{R}^{d_\rho}$, and homophily implies that nodes are unlikely to link with distant nodes. Hence, from the perspective of a given node, most other nodes are undesirable. Bounded expected degree matches the stylized fact that most social networks are *sparse* (Chandrasekhar, 2014). By comparison, some procedures for estimating network-formation models require the network to be dense, meaning that the degree of a node is order n, e.g. Graham (2014).

The increasing-domain assumption only requires nodes to be minimally different, allowing nodes to be evenly spaced apart in $\mathbb{Z}^2$, for example. From the bond percolation literature we know that this is insufficient to obtain $\mathbf{P}(i \leftrightarrow j) \to 0$. The canonical model in that literature assumes that nodes are positioned on the integer lattice and requires an extreme form of homophily in which nodes can only link with unit-adjacent nodes. In contrast, we allow for a weaker form of homophily; indeed, any pair of nodes may link with positive probability in any finite network. Hence, we instead take the approach of requiring a stronger form of dissimilarity. The next condition strengthens increasing domain and posits that larger sets of nodes are relatively more diverse or dissimilar.

Assumption 6 (Diversity)**.** *There exists $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ with $\gamma(r) < r$ satisfying the following conditions.*

(a) For any $n \in \mathbb{N}$ and $i, j \in \mathcal{N}_n$, there exists a set $\Gamma_{ij} \subseteq \mathcal{B}_i(\delta_{ij})$ such that

$$\min\{||\ell_1 - \ell_2||; \ell_1 \in \Gamma_{ij}, \ell_2 \in \{\rho_i\}_{i=1}^n \backslash \Gamma_{ij}\} \geqslant \gamma(||\delta_{ij}||).$$

(b) For $\tau(\cdot)$ defined in Assumption 4, there exists $\mu, \varphi > 0$ such that $\gamma(r)^{d_\rho}\tau(\gamma(r)) \leqslant \mu r^{-4d_\rho-\varphi}$.

Condition (a) says that if i and j are a distance r apart, then any node positioned within Γ_{ij} is at least $\gamma(r)$ apart from any node outside of this set (Figure 1). In our empirical application, we have around 120,000 PCPs dispersed across the eastern U.S., so the diversity assumption is a reasonable approximation. Combined with Assumption 5, this implies that nodes are increasingly spaced apart as $n \to \infty$. The key implication of (a) is that if i and j are connected, then there must be a linked pair of nodes on any path from i to j such that these nodes have dissimilarity at least $\gamma(||\delta_{ij}||)$. This limits the number of network paths that can possibly form between any two sufficiently dissimilar nodes, which is important for deriving a uniform bound on the probability that two nodes i and j are connected when $||\delta_{ij}||$ is large.

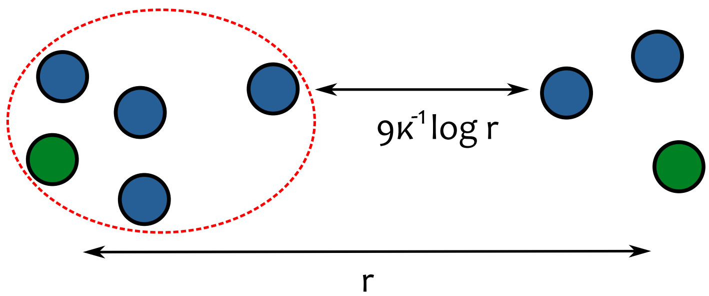


Figure 1: Visual depiction of diversity. If node i is the green node on the left and j the green node on the right, then the red circle is the j-neighborhood of i.

Remark 2 (Recursion for Diversity)**.** We show how a set of positions satisfying diversity for a given $\gamma(\cdot)$ can be generated recursively in $\mathbb{R}^2$, starting from the level of "sites," balls of diameter r_0 containing a small number of positions. This also helps to visualize sets of node positions satisfying Assumption 6. Figure 2 illustrates the recursion described below.

Step 1. Begin the recursion with one site, say, centered at the origin. This is the green site in Figure 2.

Step 2. Compute the circle of closest possible locations for a second site such that Assumptions 5 and 6(a) are satisfied. The largest possible distance between two points in each site is $2r_0 + a$, so the smallest possible distance between the sites consistent with diversity is the number a satisfying $a = \gamma(2r_0 + a)$. Then the circle of possible locations is the circumference of the ball with radius $a + \frac{r_0}{2}$ centered at the origin. Place, say, four replicas of the original site tangent to this ball, as depicted by the red sites in Figure 2. The choice of four replicas is a conservative way to ensure that all sites are at distance a away from each other.

Step 3. Treat the set of all existing sites from previous steps as a single site, and repeat step 2. This results in the blue sites, as depicted in Figure 2.

Lastly, repeat step 3 as many times as desired. The rate at which the distances a increase at each step of the recursion is implicitly determined by $\gamma(\cdot)$.

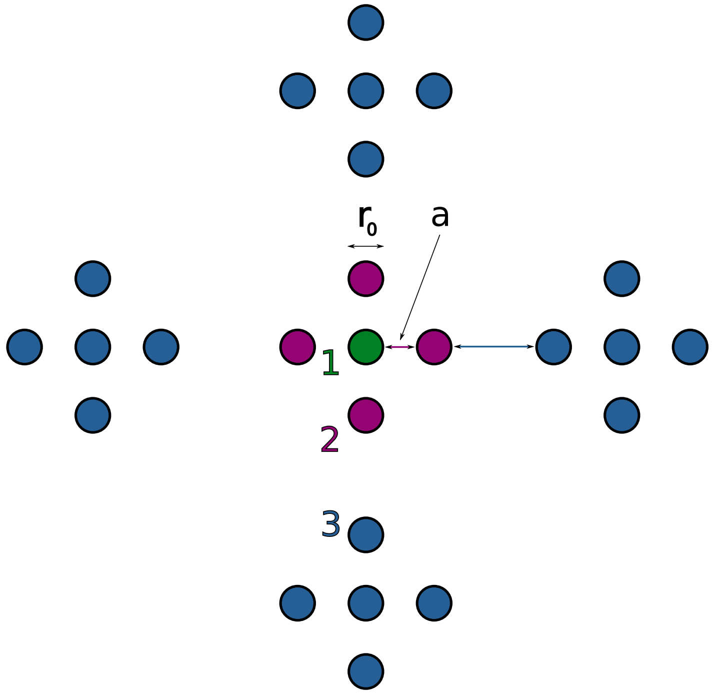


Figure 2: Diversity recursion.

Assumption 6(b) relates the tail probability τ in Assumption 4(a) to $\gamma(\cdot)$ in Assumption 6(a). This condition ensures that network dependence is sufficiently limited, i.e. that the α-mixing coefficient decays at a fast enough rate. To get a sense of how this relates to Assumption 4, which also restricts $\tau(\cdot)$ and φ, consider three cases.

- If attributes have exponential tails such that $\tau(r) = Ce^{-\kappa r}$ for some $C, \kappa > 0$, then condition (b) is satisfied with $\gamma(r) = \max\{\big((4d_\rho + \varphi + 1)\kappa^{-1}\big)\log r - \epsilon, 0\}$ for any $\epsilon \in \mathbb{R}$. Thus, if tails are thin, then $\gamma(\cdot)$ is logarithmically increasing.

- If $\tau(r) = r^{-\kappa}$ (heavy tails), then $\gamma(r)$ must be of order r^c for $c = \frac{4d_\rho+\varphi}{\kappa-d_\rho}$. If $\kappa = 5d_\rho + \varphi$, then $\gamma(r)$ is order r. Since $\gamma(r) < r$ necessarily, this implies that $\tau(r)$ cannot be of asymptotic order smaller than $r^{-5d_\rho-\varphi}$, meaning that tails cannot be too heavy for any $\gamma(\cdot)$.

- If the support of Z_{ij} bounded, then $\tau(r) = 0$ for r sufficiently large, and $\gamma(r)$ can be constant in r once r is large enough. This occurs because the latent index is uniformly bounded above in this case, and once dissimilarity exceeds that uniform bound, no links form with probability one. Hence, this is essentially the case of many independent networks.[6]

2.3 Selection Mechanism

As previously mentioned, the model is generally incomplete because for any given W, multiple networks can be consistent with (1). To complete the model, we must specify a mapping from the primitives to a particular outcome network. This is the function of the selection mechanism, which can be interpreted as a formal description of how nodes coordinate on a particular equilibrium network. The next assumption is a restriction the selection mechanism.

We need a few definitions first. Define $\mathcal{G}_{\mathcal{N}_n}(W,\theta)$ to be the correspondence that maps (W,θ) to the set of equilibrium networks on $\mathcal{N}_n$. For any $A \subseteq \mathcal{N}_n$, we call $\{G_A : G \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)\}$ the set of *equilibrium subnetworks* on A. In order to define the selection mechanism, we introduce a random vector $\nu \equiv \nu_{\mathcal{N}_n}$, unobserved by the econometrician, which functions as a *public signal* that nodes may use to coordinate on equilibrium networks. It may be depend on θ but is independent of W. As a simple example, suppose for some given W and θ there are two possible equilibrium networks. Nodes might decide to play a particular network by flipping a coin, the outcome of which we can encode in ν. Thus, ν is an unobservable that does not directly enter the latent index but affects the outcome through the selection mechanism.

Definition 2. The *selection mechanism* is a function $\lambda_{\mathcal{N}_n} : (W,\nu,\theta) \mapsto G \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)$.[7]

This says that nodes coordinate on a equilibrium network according to some function $\lambda_{\mathcal{N}_n}$, using signals W and ν, which are "common knowledge."

The set of networks $\mathcal{G}_{\mathcal{N}_n}(W,\theta)$ defines a latent partition of $\mathcal{N}_n$ such that nodes in the same partition are connected under some equilibrium network, and nodes in

[6]Strictly speaking, there can still be correlation between the unconnected components, as we clarify in the next section. In order to have full independence, Assumption 7(c) below, is also needed.

[7]The selection mechanism is more commonly defined as a mapping from W,θ to *distributions* over $\mathcal{G}_{\mathcal{N}_n}(W,\theta)$. We opt for a different definition in order to make explicit the residual randomness ν that governs selection. The implicit assumption is that ν is defined on the same probability space as W, which therefore ensures that elements of the triangular array $\{\psi_i^n; i \in \mathcal{N}_n, n \in \mathbb{N}\}$ are defined on the same probability space. This is needed in order to apply the limit results of Jenish and Prucha (2009). Furthermore, it can be shown that any distribution over $\mathcal{G}_{\mathcal{N}_n}(W,\theta)$ can be generated by some $(\lambda_{\mathcal{N}_n},\nu)$, so our construction is without loss of generality.

different partitions are not connected under any equilibrium network. It is important to note that this partition depends W and θ, since it is defined by a set of equilibrium networks. Moreover, this partition is unobserved, since W depends on unobservables. We refer to elements of this partition as *isolated societies.*

Definition 3 (Isolated Societies)**.** Let $\mathcal{S}(W,\theta)$ be the (unique) partition of $\mathcal{N}_n$ such that for any $S, S' \in \mathcal{S}(W,\theta)$ with $S \neq S'$, if $i,j \in S$ and $k \in S'$, then there exists $G \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)$ such that $i \leftrightarrow j$, and there does not exist $G \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)$ such that $i \leftrightarrow k$.

Example 2. Consider model the model

$$V_{ij}^n(G,W;\theta) = \theta_1 + \theta_2\delta_{ij} + \theta_3 G_{ji} + \varepsilon_{ij},$$

where δ_{ij} is the geographic distance between nodes i and j. Suppose there are only four nodes, with $\rho_1 = 1$, $\rho_2 = 2$, $\rho_3 = 10$, and $\rho_4 = 11$. Let $\theta = (0,-1,2)$. If $\varepsilon_{ij} = 0$ for all $i,j = 1,\dots,4$, then there are two isolated societies: $\{1,2\}$ and $\{3,4\}$. This is because there exists an equilibrium subnetwork on each of these pairs such that each node in the pair links to the other. However, due to the distance between nodes 1 and 2 and nodes 3 and 4, no node in $\{1,2\}$ ever obtains positive value from linking to a node in $\{3,4\}$, so the two societies are isolated.

Now suppose that $\varepsilon_{ij} > 7$ if $i \in \{1,2\}$ and $j \in \{3,4\}$ and $\varepsilon_{ij} = 0$ otherwise. Then all four nodes are part of the same isolated society, since it is an equilibrium for pairs $(1,3), (2,4), (1,2)$, and $(3,4)$ to be linked bidirectionally.

Due to component externalities (Assumption 3), the subnetwork formed by nodes in a society S does not enter the latent index of nodes in another society S' because the two are not connected in equilibrium. Intuitively, then, we might expect that nodes in different societies should form subnetworks independently. However, there exist selection mechanisms such that S and S' still "coordinate" on the subnetworks they form, despite being isolated. That is, the equilibrium subnetworks selected by different societies may be statistically dependent. To see this, consider a six-node model in which, for some fixed W, two triplets form isolated societies, which we will call "blue" and "red." Further suppose that each isolated society has two possible equilibrium subnetworks: the fully connected subnetwork and the empty subnetwork.

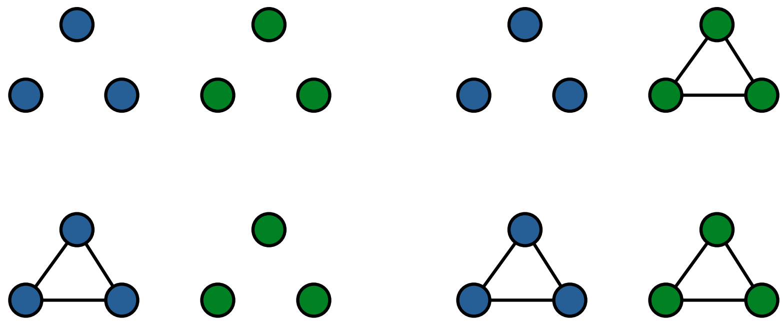


Figure 3: Global coordination.

Then there are four possible equilibrium networks in the overall network, displayed in Figure 3. Consider the selection mechanism that selects the network in the top-left panel with probability $\frac{1}{2}$ and the network in the bottom-right panel with probability $\frac{1}{2}$. Then the equilibrium subnetworks selected by the two societies are perfectly correlated.

The next assumption rules out this type of coordination, requiring that the selection mechanism chooses from $\mathcal{G}_{\mathcal{N}_n}(W,\theta)$ by *independently* randomizing over equilibrium subnetworks in $\mathcal{G}_S(W_S,\theta)$ for each $S \in \mathcal{S}(W,\theta)$. (Note that $\mathcal{G}_S(W_S,\theta)$ is the set of equilibrium networks on S if the only nodes present were those in S.) Independent selection is an intuitive requirement, since (a) isolated societies are unconnected, and (b) Assumption 3 implies that the subnetworks formed by unconnected nodes are not payoff-relevant, and therefore, isolated societies have no incentive to coordinate.

We need one last piece of notation. For $S \in \mathcal{S}(W,\theta)$, let $\lambda_{\mathcal{N}_n}(W,\nu,\theta)\big|_S$ be the restriction of the range of $\lambda_{\mathcal{N}_n}$ to $\mathbf{G}_S$. This restricted range equals $\{G_S : G \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)\}$, which by Assumption 3 equals $\mathcal{G}_S(W_S,\theta)$.

Assumption 7 (Selection Mechanism)**.** *There exist a sequence of selection mechanisms $\{\lambda_{\mathcal{N}_n}\}_{n=2}^{\infty}$ and random vectors $\{\nu_{\mathcal{N}_n}\}_{n=2}^{\infty}$ such that, for n sufficiently large, the following hold with probability one.*

(a) (Coherence) $|\mathcal{G}_{\mathcal{N}_n}(W,\theta_0)| \geqslant 1$.

(b) (Rationalizability) For $g \in \mathbf{G}_n$,

$$\mathbf{P}(G = g \,|\, W) = \mathbf{P}(\lambda_{\mathcal{N}_n}(W,\nu_{\mathcal{N}_n},\theta_0) = g \,|\, W).$$

(c) *(No Coordination) There exist selection mechanisms* $\{\lambda_S; S \in \mathcal{S}(W, \theta_0)\}$ *and independent random variables* $\{\nu_k\}_{k=1}^{|\mathcal{S}(W,\theta_0)|}$ *independent of* W*, such that for all* $S \in \mathcal{S}(W, \theta_0)$

$$\lambda_{\mathcal{N}_n}(W, \nu_{\mathcal{N}_n}, \theta_0)\big|_S = \lambda_S(W_S, \nu_S, \theta_0).$$

Part (a) ensures that, under θ_0, each W is mapped to some G, meaning that the set of equilibrium networks is nonempty, which guarantees that ψ_i^n is well defined. Sheng (2014) provides sufficient conditions for coherence. Condition (b) defines the selection mechanism, requiring that the true data generating process is rationalized by model (1) and some selection mechanism. Condition (c) restricts the sequence of selection mechanisms over n, requiring that, for each W, isolated societies choose their subnetworks according to independent subnetwork selection mechanisms. Furthermore, the independence restriction on the society-specific signals ν_S ensures that isolated societies do not coordinate in the sense that the realizations of their respective subnetworks are independent. Thus, if $i \notin S \in \mathcal{S}(W, \theta)$, then members of S play a particular subnetwork irrespective of the realizations of the subnetwork and the attributes of nodes in i's society. Note that it is well defined to select the entire network by selecting subnetworks on isolated societies, as dictated by (c), since $\mathcal{G}_S(W_S, \theta_0)$ is the range of $\lambda_S(W_S, \nu_S, \theta_0)$, and as previously argued, this is equivalent to the range of $\lambda_{\mathcal{N}_n}(W, \nu, \theta_0)\big|_S$.

It is useful to compare (c) with the standard independent sampling framework in which the econometrician observes many independent networks. We can view the union of these networks as a single large network. In our framework, the underlying independent subnetworks are the isolated societies $\mathcal{S}(W, \theta)$, which are *unknown*, since W contains unobservables. Thus, assuming that two observed subnetworks are independent, as in the standard many-markets framework, requires additional knowledge about the isolated societies. Specifically, the econometrician must posit that (a) nodes in the two subnetworks can never be linked to each other, and (b) nodes in these subnetworks do not coordinate. These assumptions imply that under many-market asymptotics, the econometrician has knowledge of a potentially coarser partition of nodes than $\mathcal{S}(W, \theta)$, such that the elements of this partition are unions of isolated societies. In contrast, our approach requires no knowledge of $\mathcal{S}(W, \theta)$. In practice, this eliminates the need in the many-markets framework to manually partition the observed network into many independent subnetworks in order to generate multiple "observations."

2.4 Main Results

We can now state the core theorem of the paper.

Theorem 1. *Under Assumptions 1-7, $\{\psi_i^n; i \in \mathcal{N}_n, n \in \mathbb{N}\}$ is an α-mixing random field.*

The intuition behind the result is that as the dissimilarity between nodes i and j grows,

(i) the probability that they are connected is small, and

(ii) if they are unconnected, their referral decisions, and thus their node statistics, should be independent.

Hence, we should expect that the mixing coefficient vanishes as the dissimilarity diverges. This intuition can be formalized by decomposing the mixing coefficient into two components. Using the law of total probability, it is not difficult to show that

$$\bar{\alpha}_{1,1}\big(||\delta_{ij}||\big) \leqslant 3\mathbf{P}(i \leftrightarrow j) + \bar{\alpha}_{1,1}\big(||\delta_{ij}|| \;\big|\, i \nleftrightarrow j\big), \tag{3}$$

where $\bar{\alpha}_{1,1}\big(||\delta_{ij}|| \;\big|\, i \nleftrightarrow j\big)$ is the conditional mixing coefficient with the unconditional probabilities in the definition replaced with probabilities conditional on the event that the two nodes are disconnected. This captures the extent to which isolated societies (disconnected sets of nodes) "coordinate." To show that the mixing coefficieent vanishes, it suffices to show that the two elements on the right of (3) decay to zero. The intuition of (i) corresponds to $\mathbf{P}(i \leftrightarrow j) \to 0$ and is ensured by homophily (Assumption 4), which states that the probability a pair of nodes is connected decreases as the nodes become more dissimilar, and diversity (Assumption 6), which ensures sufficient dissimilarity in the network. The intuition of (ii) corresponds to setting $\bar{\alpha}_{1,1}\big(||\delta_{ij}|| \;\big|\, i \nleftrightarrow j\big) = 0$, which is ensured by component externalities (Assumption 3) and no coordination (Assumption 7(c)). These imply that network components are independent, conditional on the nodes being unconnected. The formal proof shows that the probability that two nodes are connected is vanishing at a sufficiently fast rate in the distance between the nodes. The tail condition (Assumption 4(a)) and the rate condition (Assumption 6(b)) play crucial roles here.

To understand why diversity is important to ensure that $\mathbf{P}(i \leftrightarrow j) \to 0$ as $\delta_{ij} \to \infty$, notice that the connection probability is the union of an extremely large number of events:

$$\mathbf{P}(i \leftrightarrow j) = \mathbf{P}\left(\underbrace{\bigcup_{\ell=1}^{n-1}}_{\text{path length}} \underbrace{\bigcup_{\substack{k_0,\dots,k_\ell \in \mathcal{N}_n: \\ k_0=i, k_\ell=j}}}_{\text{nodes in path}} \left\{ G_{k_0,k_1} = \cdots = G_{k_{\ell-1},k_\ell} = 1 \right\} \right). \tag{4}$$

That is, there are many possible paths that can connect a pair of nodes. For example, for 10 nodes, there are 1012 possible paths connecting any two nodes. We cannot bound this probability using the usual the union bound, since the number of events in this union is too large, more than exponential in n. Diversity allows us to significantly reduce the number of events in this union in order to apply the union bound. To see this, let $\Gamma_{ij} = \{k \in \mathcal{N}_n : ||\delta_{ik}|| \leqslant ||\delta_{ij}||\}$. Diversity implies that if nodes i and j are connected, then some pair of directly linked nodes (k, l) along that chain, with $k \in \Gamma_{ij}$, must be approximately $\gamma\big(\delta_{ij}||\big)$ apart. Thus,

$$(4) \leqslant \mathbf{P}\left(\bigcup_{\substack{k,l:\, k \in \Gamma_{ij}, \\ \delta_{kl} \geqslant \gamma\left(\delta_{ij}||\right)}} \left\{ G_{kl} = 1 \right\} \right). \tag{5}$$

This significantly reduces the number of events in the union. We can then apply the union bound:

$$\begin{aligned}(5) &\leqslant \sum_{k \in \Gamma_{ij}} \sum_{l=1}^{n} \mathbf{P}\left(G_{kl} = 1 \,|\, \delta_{kl} \geqslant \gamma\big(\delta_{ij}||\big)\right) \\ &\leqslant \big(\#k \in \Gamma_{ij}\big) \max_k \sum_{l=1}^{\infty} \mathbf{P}\left(G_{kl} = 1 \,|\, \delta_{kl} \geqslant \gamma\big(\delta_{ij}||\big)\right). \end{aligned} \tag{6}$$

It remains to show that the right-hand side of the last expression converges to zero as $||\delta_{ij}|| \to \infty$. Intuitively, for a fixed node k, as the sum over l ranges over nodes increasingly distant from k, the summands on the RHS should be progressively decaying to zero. This is because, by homophily and the tails condition, G_{kl} is unlikely to equal one if k and l are far apart. Hence, the series remains finite for any $||\delta_{ij}||$. Furthermore, as $||\delta_{ij}|| \to \infty$, the conditions ensure that the entire summation decays to zero by homophily, and at a rate faster than the growth of the set Γ_{ij}. Thus (6) is shrinking to zero, ensuring that $\mathbf{P}(i \leftrightarrow j) \to 0$ as $||\delta_{ij}|| \to \infty$.

Having established α-mixing, a law of large numbers and central limit theorem then follow from results in the spatial econometrics literature.

Theorem 2. *Suppose that $\{\psi_i^n; i \in \mathcal{N}_n, n \in \mathbb{N}\}$ is an α-mixing random field and that there exists an array of positive real constants $\{c_i^n\}$ such that*

(a) $\lim_{k\to\infty} \sup_n \max_{i\in\mathcal{N}_n} \mathbf{E}\left[\left|\frac{\psi_i^n}{c_i^n}\right|^3 \mathbf{1}\left\{\left|\frac{\psi_i^n}{c_i^n}\right| > k\right\}\right] = 0,$ *and*

(b) $\liminf_{n\to\infty} \sigma_n^2/(nM_n^2) > 0,$ *where* $M_n = \max_i c_i^n$ *and* $\sigma_n^2 = \text{Var}\left(\sum_i \psi_i^n\right)$.

Then $\frac{1}{nM_n}\sum_i(\psi_i^n - \mathbf{E}\psi_i^n) \xrightarrow{L_1} 0$ *and* $\sigma_n^{-1}\sum_i \psi_i^n \xrightarrow{d} N(0,1)$.

Assumptions (a) and (b) are sufficient for a Lindeberg condition. The scaling constants $\{c_i^n\}$ allow for asymptotically unbounded moments. If node statistics are uniformly bounded, then we can take $c_i^n = 1$. Otherwise, we generally choose $c_i^n = \sqrt{\mathbf{E}[(\psi_i^n)^2]}$.

Remark 3. The class of network statistics for which Theorem 2 holds is quite general and includes complex statistics such as the average clustering coefficient and average path length. Define

$$Cl_i(G) = \frac{\sum_{j\neq i;k\neq j;k\neq i} G_{ij}G_{ik}G_{jk}}{\sum_{j\neq i;k\neq j;k\neq i} G_{ij}G_{ik}}.$$

This is the proportion of nodes link to i that are also linked to each other, with the convention that $Cl_i(G) := 0$ if i has at most one link. The *average clustering coefficient* of G is $\frac{1}{n}\sum_{i=1}^n Cl_i(G)$. Here $\psi_i^n = Cl_i(G)$, and $M_n = 1$, since $Cl_i(G) \in [0,1]$.

Let $Pl_{ij}(G)$ be the length of the shortest path between nodes i and j. The *average path length* of G is then $\frac{1}{nM_n}\sum_{i=1}^{n-1}\sum_{j=i+1}^n Pl_{ij}(G)$. Here $\psi_i^n = \sum_{j=i+1}^n Pl_{ij}(G)$.

Theorem 2 is useful because it enables us to consistently estimate subnetwork moments (expectations of network statistics), and certain moments can be used to

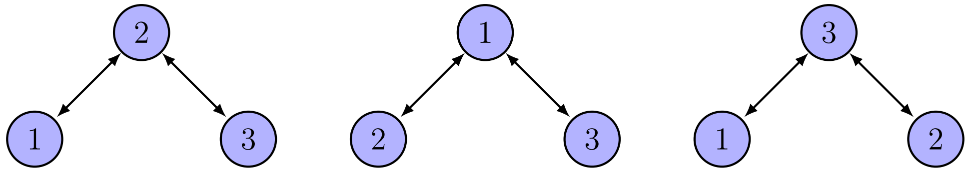


Figure 4: Equivalence class of intransitive triads.

define an identified set for θ_0. Before presenting these moments, we need a few definitions. Two labeled subnetworks $g_A, g'_A \in \mathbf{G}_A$ are *isomorphic* (denoted $g_A \sim g'_A$) if there exists a permutation of the indices in A, denoted $\sigma(A)$, such that $g_A = g'_{\sigma(A)}$. The equivalence class of g_A is $\{g_S \in \mathbf{G}_S : S \subseteq \mathcal{N}_n, g_A \sim g_S\}$. For example, in Figure 4, we have three labeled subnetworks isomorphic to one another that together constitute an equivalence class known as the *intransitive triads*.

The set of unlabeled subnetworks of size a, denoted $\mathbf{G}_{u,a}$, is the set of equivalence classes of subnetworks of size a. For example, $\mathbf{G}_{u,3}$ consists of four equivalence classes: the empty network, the network with only one link, the network with only two links (intransitive triads), and the fully connected network (transitive triads). We are interested in the subset of $\mathbf{G}_{u,a}$ consisting only of *connected* subnetworks, which we denote by $\mathbf{G}^{\leftrightarrow}_{u,a}$. Relative to $\mathbf{G}_{u,a}$, this set does not contain subnetworks with isolated nodes. Thus, for $a = 3$, $\mathbf{G}^{\leftrightarrow}_{u,a}$ only consists of two elements: the intransitive triads and the transitive triads.

Let $g_a \in \mathbf{G}^{\leftrightarrow}_{u,a}$. Then for $|A| = a$, $\mathbf{P}(G_A \in g_a)$ is the probability that the set of nodes on A forms a subnetwork in the equivalence class g_a. For example, if g_3 is the equivalence class of intransitive triads, $\mathbf{P}(G_{\{i,j,k\}} \in g_3) = \mathbf{P}(G_{ij}G_{jk}(1 - G_{ik}) = 1) + \mathbf{P}(G_{jk}G_{ik}(1 - G_{ij}) = 1) + \mathbf{P}(G_{ik}G_{ij}(1 - G_{jk}) = 1)$. Intuitively, moments such as the probability that triplets form intransitive triads and the probability that they form transitive triads are useful to include if we are interested in, say, identifying θ_4 in Example 1

We can estimate averages of probabilities $\mathbf{P}(G_A \in g_a)$ over subsets of nodes $A \subseteq \mathcal{N}_n$ with the same cardinality $|A| = a$ for *connected* unlabeled subnetworks g_a. By Theorem 2,

$$\frac{1}{nM_n} \sum_{\substack{A \subseteq \mathcal{N}_n: \\ |A| = a}} \left(\mathbf{1}\{G_A \in g_a\} - \mathbf{P}\left(G_A \in g_a\right)\right) \xrightarrow{p} 0,$$

for any $g_a \in \mathbf{G}^{\leftrightarrow}_{u,a}$.[8] To see this, notice that

$$\frac{1}{nM_n} \sum_{\substack{A \subseteq \mathcal{N}_n: \\ |A|=a}} \mathbf{1}\{G_A \in g_a\} = \frac{1}{nM_n} \sum_{\substack{A \subseteq \mathcal{N}_n: \\ |A|=a}} \sum_{g_A \in g_a} \mathbf{1}\{G_A = g_A\}$$

$$= \frac{1}{anM_n} \sum_{i=1}^n \underbrace{\sum_{\substack{A \subseteq \mathcal{N}_n: i \notin A, \\ |A|=a-1}} \sum_{g_{A\cup\{i\}} \in g_a} \mathbf{1}\{G_{A\cup\{i\}} = g_{A\cup\{i\}}\}}_{\psi_i^n}.$$

In order for ψ_i^n to be a valid node statistic, g_a must be connected, since this ensures that ψ_i^n only depends on the component of node i. This means, for example, that if $A = \{1,2,3\}$, then g_a cannot be isomorphic to the empty subnetwork, or the subnetwork in which $G_{12} = 1$ and the other potential links are zero, since node 3 is isolated.

3 Inference

With Theorem 2, we can estimate subnetwork moments. We next describe how to leverage these moments for inference on θ_0. Generally, model (1) admits multiple equilibrium networks, meaning that $\mathcal{G}_{\mathcal{N}_n}(W, \theta_0)$ is non-singleton. Aside from Assumption 7(c), theory typically imposes few restrictions on the selection mechanism. Thus, we seek to derive moment inequalities that enable inference on θ_0 without imposing additional assumptions on the selection mechanism. The inference procedure will require some additional conditions.

Let $Z_A^o = \big((X_i^o, X_j^o, \zeta_{ij}^o); i,j \in A\big)$ and $\varepsilon_A = \big((X_i^u, X_j^u, \zeta_{ij}^u); i,j \in A\big)$.

Assumption 8 (Distribution). *For any $A \subseteq \mathcal{N}_n$, the conditional distribution of $\varepsilon_A \,|\, Z_A^o$ is continuous and known up to a finite-dimensional parameter that does not depend on A. Without loss of generality, this parameter is a subvector of θ_0. Additionally, for each $i,j \in A$, $V_{ij}^n(\cdot)$ is continuous in ε_A.*

This assumption is standard in the literature (e.g. Boucher and Mourifié, 2013; Mele, 2013; Sheng, 2014). The next assumption is important for ensuring the computational

[8]Proposition B.1 shows that in our model we can take the scaling constant M_n to be equal to one when attributes have exponential tails.

feasibility of the inference procedure.

Assumption 9 (Endogenous Factors)**.**

(a) There exists $\xi \in \mathbb{N}$ known to the econometrician such that for all n and $i, j \in \mathcal{N}_n$, the support of $\mathcal{E}_{ij}^n$ has cardinality ξ.

(b) For any n and $i, j \in \mathcal{N}_n$, $\mathcal{E}_{ij}^n(G, W) = \mathcal{E}_{ij}^n(G, Z_{ij})$, where $Z_{ij} = (X_i, X_j, \zeta_{ij})$.

Condition (a) is a restriction on network externalities that substantially weakens Sheng's (2014) local externalities assumption under a capacity constraint, as we discuss further below. While (a) may seem strong, in fact it is satisfied by our previous examples in Section 2. When $\mathcal{E}_{ij}^n$ does not depend on Z_{ij}, (a) is often a byproduct of requiring uniform boundedness (Assumption 2(b)). On the other hand, if the endogenous factors do depend on Z_{ij}, then we need to restrict its support. Condition (b) is imposed for technical reasons, as it ensures that the average of certain simulated moments constitutes a finite-order U-statistic. Sheng also imposes this requirement. It is likely feasible to dispense with (b) using the theory of infinite-order U-statistics (Frees, 1989), but this complicates the analysis in Proposition B.2. We leave this to future work.

3.1 Moment Inequalities

In this section, we define the moment inequalities used for inference. These inequalities are valid without having to impose the assumptions made in Section 2, which are used to establish α-mixing, with the exception of Assumption 7(b). This assumption states that the model model (1), which imposes a stability requirement on the link formation process, rationalizes the data. That is the pair of nodes (i, j) "best-responds" to the rest of the network G by choosing G_{ij} to maximize the joint surplus $V_{ij}^n(G, W; \theta_0)$. By leveraging the implications of the stability requirement and variation in linking frequencies between sets of nodes with different attributes, the moment inequalities we derive will partially identify θ_0.

First we need some notation. Let $A \subseteq \mathcal{N}_n$, $a = |A|$, and $d_a = |\mathbf{G}_{u,a}^{\leftrightarrow}|$. We normalize $||\theta|| \leqslant 1$ for all $\theta \in \Theta$. Recall that, for any $g \in \mathbf{G}_n$, $g_{-A} \equiv \{g_{ij} : i, j \in \mathcal{N}_n, \text{ not both in } A\}$.

By Assumption 7(b), for $\theta = \theta_0$, there exists a sequence of selection mechanisms $\{\lambda_{\mathcal{N}_n}\}$ such that for all $n \in \mathbb{N}$ and $A \subseteq \mathcal{N}_n$,

$$\mathbf{P}(G_A = g_A \,|\, Z_A^o) = \sum_{g_{-A}} \mathbf{E}_\theta \left[\mathbf{1}\left\{(g_A, g_{-A}) \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)\right\} \right. \\ \left. \times \mathbf{1}\left\{\lambda_{\mathcal{N}_n}(W,\nu,\theta) = (g_A, g_{-A})\right\} \,\middle|\, Z_A^o \right] \quad (7)$$

with probability one. We use this to derive a set of moment inequalities that are valid at the identified set of parameters. We first convert the conditional moment (7) to a set of unconditional moments using Andrews and Shi (2013) instruments. Specifically, θ satisfies (7) if and only if it satisfies

$$\mathbf{E}[\mathbf{1}\{G_A = g_A\} f_a(Z_A^o)] = \sum_{g_{-A}} \mathbf{E}_\theta \left[\mathbf{1}\left\{(g_A, g_{-A}) \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)\right\} \right. \\ \left. \times \mathbf{1}\left\{\lambda_{\mathcal{N}_n}(W,\nu,\theta) = (g_A, g_{-A})\right\} f_a(Z_A^o) \right] \quad \forall f_a \in \mathcal{F}_a, \quad (8)$$

where $\mathcal{F}_a$ is the set of functions of Z_A^o (instruments) $f_a : \mathbb{R}^{ad_x} \to \{0,1\}$ such that f_a is symmetric in Z_A^o.[9] The "only if" direction is clear. The "if" direction follows since $\mathcal{F}_a$ contains the contains the class of "countable hypercube" instruments (see Example 1 of Andrews and Shi (2013)), which implies (7) by Lemmas 2 and 3 of Andrews and Shi (2013). An example of f_a is the indicator for whether or not nodes in A all have the same first attribute.

We can consistently estimate the average of the left-hand side directly from the data for all connected, unlabeled subnetworks $g_a \in \mathbf{G}_{u,a}^{\leftrightarrow}$. By Theorem 2,

$$\lim_{n\to\infty} \frac{1}{nM_n} \sum_{\substack{A \subseteq N_n: \\ |A|=a}} \sum_{g_A \in g_a} \big(\mathbf{1}\{G_A = g_A\} f_a(Z_A^o) - \mathbf{E}\left[\mathbf{1}\{G_A = g_A\} f_a(Z_A^o)\right] \big) = 0.$$

Consequently, in light of (8), we define the *identified set* as follows.

[9]Symmetry is a natural requirement, since these moments involve unlabeled networks. It is also needed to apply a law of large numbers for U-statistics.

Definition 4. The identified set is

$$\Theta_I = \Bigg\{\theta \in \Theta : \exists \{\lambda_{\mathcal{N}_n}\} \text{ satisfying Assumption 7 such that}$$

$$\lim_{n\to\infty} \frac{1}{nM_n} \sum_{\substack{A\subseteq\mathcal{N}_n:\\ |A|=a}} \sum_{g_A\in g_a} \Bigg(\underbrace{\mathbf{E}\left[\mathbf{1}\{G_A = g_A\} f_a(Z_A^o)\right]}_{\mathbf{I}} -$$

$$\underbrace{\sum_{g_{-A}} \mathbf{E}_\theta \left[\mathbf{1}\big\{\lambda_{\mathcal{N}_n}(W,\nu,\theta) = (g_A, g_{-A})\big\} f_a(Z_A^o)\right]}_{\mathbf{II}} \Bigg) = 0$$

$$\text{for all } f_a \in \mathcal{F}_a, g_a \in \mathbf{G}_{u,a}^{\leftrightarrow}, a > 1\Bigg\}.$$

The average of component **I** is our subnetwork moment, which can be estimated directly from the data, while **II** is the model moment. Then Θ_I is the set of parameters that match these two moments. This is the analog of the standard definition of the sharp identified set in a many-markets context (see e.g. Beresteanu et al., 2011; Galichon and Henry, 2011; Tamer, 2010). The difference is that in the large-market context, we do not observe the joint distribution $\mathbf{P}(G = g \mid X^o)$ but instead its observe analog for subsets of nodes, namely subnetwork moments.

Because Θ_I depends on a sequence of unknown functions $\lambda_{\mathcal{N}_n}$, it is not immediately useful for inference. We next provide a more practical characterization of an identified set that removes the dependence on the nuisance parameters $\lambda_{\mathcal{N}_n}$. Arbitrarily label the elements of $\mathbf{G}_{u,a}^{\leftrightarrow}$ as $g_a^1, \dots, g_a^{d_a}$. For $g_A \in \mathbf{G}_A$, define

$$\mathbb{I}_a(g_A) = \left(\mathbf{1}\{g_A \in g_a^1\}, \dots, \mathbf{1}\{g_A \in g_a^{d_a}\}\right).$$

Note that this can be a vector of zeros. Further, let $\mathcal{G}_A(g_{-A}, W_A, \theta)$ be the set of subnetworks $g_A \in \mathbf{G}_A$ such that for $i, j \in A$, we have $g_{ij} = 1$ if and only if $V_{ij}\left((g_A, g_{-A}), W_A, \theta\right) \geqslant 0$.[10] In other words, this is the set of subnetworks on A that

[10]Note that $V_{ij}(\cdot)$ only depends on W through W_A by Assumption 9.

are equilibrium subnetworks under g_{-A}. Let $\mathcal{U}_a = \{u \in \mathbb{R}^{d_a} : ||u|| \leqslant 1\}$, and define

$$\bar{m}_a^n(\theta; u, f_a) = \frac{1}{nM_n} \sum_{\substack{A \subseteq \mathcal{N}_n: \\ |A|=a}} \Bigg(u'\mathbb{I}_a(G_A) f_a(Z_A^o) \\ - \mathbf{E}_\theta \left[\max_{g_{-A}} \max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)} u'\mathbb{I}_a(g_A) f_a(Z_A^o) \,\middle|\, Z_A^o \right] \Bigg)$$

for $u \in \mathcal{U}_a$. Notice the conditional expectation is taken under θ with respect to W_A, which contains unobservables.

Definition 5. The *(computable) identified set* is

$$\Theta_{IC} = \left\{ \theta \in \Theta : \limsup_{n\to\infty} \mathbf{E}_\theta \left[\bar{m}_a^n(\theta; u, f_a) \right] \leqslant 0, \forall u \in \mathcal{U}_a, f_a \in \mathcal{F}_a, a \in \{2, \dots, \bar{a}\} \right\}. \quad (9)$$

We discuss $\bar{a}$ below. The term $\frac{1}{n}\sum_{\substack{A \subseteq \mathcal{N}_n: \\ |A|=a}} u'\mathbf{E}\big[\mathbb{I}_a(G_A) f_a(Z_A^o)\big]$ in $\mathbf{E}\left[\bar{m}_a^n(\theta; u, f_a)\right]$ is a convex combination of the moments $\frac{1}{n}\sum_{\substack{A \subseteq \mathcal{N}_n: \\ |A|=a}} \sum_{g_A \in g_a} \mathbf{E}\left[\mathbf{1}\{G_A = g_A\} f_a(Z_A^o)\right]$ that appear in Θ_I. The main difference between Θ_I and Θ_{IC} lies in the second terms. Essentially, we take a convex combination of term **II** in Θ_I and replace this with an upper bound

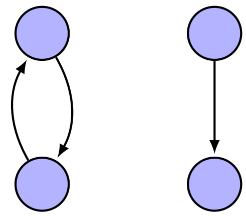
$$\mathbf{E}_\theta \left[\max_{g_{-A}} \max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)} u'\mathbb{I}_a(g_A) f_a(Z_A^o) \right]$$

that does not depend on this nuisance parameter, converting the moment equality in Θ_I to a moment inequality in Θ_{IC}.

Example 3. Suppose G is a directed network. We illustrate the construction of $\max_{g_{-A}} \max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)} u'\mathbb{I}_a(g_A)$ for $|A| = 2$ for the model

$$V_{ij}(G, W; \theta) = Z_{ij}'\theta_z + G_{ji}\theta_r + T_{ij}\theta_t,$$

where $T_{ij} = \mathbf{1}\{\exists\, k : G_{ik} = G_{kj} = 1\}$. (Recall $Z_{ij} = (X_i, X_j, \zeta_{ij})$.) Let g_2^b be the equivalence class of linked pairs of nodes with bidirectional linking and g_2^d unidirectional linking (Figure 5). For $|A| = 2$, $\mathbb{I}_a(G_A) = \big(\mathbf{1}\{G_A \in g_2^b\}, \mathbf{1}\{G_A \in g_2^d\}\big)$.



(a) g_2^b (b) g_2^d

Figure 5: Equivalence classes of dyads.

Define the events

$$\begin{aligned}\mathcal{B} &= \left\{Z_{ij}'\theta_z + \theta_r + T_{ij}\theta_t \geqslant 0 \cap Z_{ji}'\theta_z + \theta_r + T_{ji}\theta_t \geqslant 0\right\}, \\ \mathcal{D} &= \left\{Z_{ij}'\theta_z + T_{ij}\theta_t \geqslant 0 \cap Z_{ji}'\theta_z + T_{ji}\theta_t < 0\right\} \\ &\qquad\qquad \cup \left\{Z_{ij}'\theta_z + T_{ij}\theta_t < 0 \cap Z_{ji}'\theta_z + T_{ji}\theta_t \geqslant 0\right\}.\end{aligned}$$

To interpret these events, fix g_{-A} (thus fixing T_{ij} and T_{ji}). Then $\mathcal{B}$ is the event, according to the model, that bidirectional linking is an equilibrium on the dyad consisting of nodes i and j under g_{-A}, while $\mathcal{D}$ is the event that unidirectional linking is an equilibrium under g_{-A}. Let $u = (u^b, u^d)$, where u^b is associated with g_2^b and u^d is associated with g_2^d. One can show that

$$\max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)} u'\mathbb{I}_a(g_A) = \max\{u^b, u^d\}\mathbf{1}\{\mathcal{B} \cap \mathcal{D}\} + u^b\mathbf{1}\{\mathcal{B} \cap \mathcal{D}^c\} + u^d\mathbf{1}\{\mathcal{D} \cap \mathcal{B}^c\},$$

where $\mathcal{B}^c$ denotes the complement of the event $\mathcal{B}$. This can be seen by considering the three cases in which $\mathcal{G}_A(g_{-A}, W_A, \theta)$ contains networks isomorphic to both g_2^b and g_2^d, or just one or the other. Consequently,

$$\begin{aligned}&\max_{g_{-A}} \max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)} u'\mathbb{I}_a(g_A) \\ &\qquad = \max_{T_{ij}, T_{ji} \in \{0,1\}} \left\{\max\{u^b, u^d\}\mathbf{1}\{\mathcal{B} \cap \mathcal{D}\} + u^b\mathbf{1}\{\mathcal{B} \cap \mathcal{D}^c\} + u^d\mathbf{1}\{\mathcal{D} \cap \mathcal{B}^c\}\right\},\end{aligned}$$

since the maximum over g_{-A} on the left-hand side corresponds to the maximum over T_{ij}, T_{ji} on the right-hand side.

We next establish the relationship between the computable identified set and the identified set.

Proposition 3.1. *Under Assumptions 2, 7(b), and 9(a),* $\Theta_{IC} \supseteq \Theta_I$.

Hence, Θ_{IC} is conservative. We next argue that (A) conservative inference is the cost of computational feasibility in our setting and that, nonetheless, (B) the set is sharper than feasible alternatives in the existing literature. Regarding (A), our computable identified set is closely related to the characterization of the identified set for games of complete information due to Beresteanu et al. (2011), who study the case in which the econometrician observes many markets. The analog of their set in the large-market case is similar to Θ_{IC}, except we would replace $\max_{g_{-A}} \max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)}$ in the definition of $\mathbf{E}\left[\bar{m}_a^n(\theta; u, f_a)\right]$ with $\max_{g \in \mathcal{G}_{\mathcal{N}_n}(W,\theta)}$.[11] Our set differs for computational reasons. In general, the conditional expectation in (9) must be computed by simulation, and computing $\mathcal{G}_{\mathcal{N}_n}(W,\theta)$ is computationally infeasible, since the number of equilibrium networks typically grows exponentially with n. In order to avoid this curse of dimensionality, we settle for computing a more conservative set by taking the max over a larger set. For a given g_{-A} the maximum over $g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)$ is feasible, at least if $|A|$ is not too large. This is why we restrict $a \in [2, \bar{a}]$, with $\bar{a}$ chosen according to computational capabilities, keeping in mind the size of $\mathbb{R}^{d_a}$. For example, Sheng (2014) chooses $\bar{a} = 5$. The maximum over g_{-A} turns out to be feasible under Assumption 9, since $V_{ij}(\cdot)$ defines $\mathcal{G}_A(g_{-A}, W_A, \theta)$, and for fixed W and θ, the number of possible distinct values of $V_{ij}(\cdot)$ is no more than ξ. Therefore this maximum is taken over a finite number of values, which are known. For instance, in Example 3, the maximum is taken over only four values, namely the possible values of (T_{ij}, T_{ji}), where T_{ij} and T_{ji} are binary.

Regarding (B), there are two papers in the existing literature that develop moment inequalities for inference in the many-networks case, but with Theorem 2, in theory, these can be used in the large-network setting. Both papers develop conservative bounds, as well. The first set of bounds, developed by Miyauchi (2013), are computationally infeasible for a large-network setting; because they require computing equilibrium networks, they suffer from a curse of dimensionality.[12] Sheng (2014)

[11] It is an open question whether or not this replacement yields a set that is equivalent to Θ_I in the large-market case.

[12] Miyauchi's approach also requires a strong restriction on preferences ("non-negative externalities"), and the network statistics used to construct his moments must also satisfy a monotonicity

develops Ciliberto-Tamer-type bounds that are feasible in a large-network setting.[13] Our bounds are sharper, since we attain the analog of Sheng's bounds by restricting all $u \in \mathcal{U}_a$ to be the canonical basis vectors multiplied by ± 1 (cf. Beresteanu et al., 2011, Proposition 3.3). Furthermore, our Assumptions 3 and 9(a) substantially weaken her local externalities assumption, which restricts $\mathcal{E}_{ij}^n$ to only depend on nodes at most two links away from i or j. In contrast, our assumptions allow $V_{ij}(\cdot)$ to depend on nodes any finite number of links away from i or j.

3.2 Inference Procedure

Next, we discuss the inference procedure for θ_0. Let $A+i = A \cup \{i\}$. Define the moment function

$$m_{i,a}^n(\theta; u, f_a) = \sum_{\substack{A \subseteq \mathcal{N}_n : i \notin A \\ |A| = a-1}} \Bigg(u'\mathbb{I}_a(G_{A+i}) f_a(Z_{A+i}^o) \\ - \mathbf{E}_\theta \left[\max_{g_{-(A+i)}} \max_{\substack{g_{A+i} \in \\ \mathcal{G}_{A+i}(g_{-(A+i)}, \theta)}} u'\mathbb{I}_a(g_{A+i}) f_a(Z_{A+i}^o) \,\middle|\, Z_{A+i}^o \right] \Bigg). \quad (10)$$

where $u \in \mathcal{U}_a$. This is the difference between two components. As previously noted, the first component is a node statistic, so its average over i is consistent for its expectation by Theorem 2. The second component is a function of Z_A^o, so its average over all subsets of nodes A of size a constitutes a U-statistic of order a. Consistency for its expectation follows by Proposition B.2 and a standard law of large numbers. This component typically lacks a closed-form expression. However, it can be simulated, since the conditional density of ε_A given Z_A^o is known (Assumption 8).

The empirical moments are $\bar{m}_a^n(\theta; u, f_a) = \frac{1}{n d_a M_n} \sum_{i=1}^n m_{i,a}^n(\theta; u, f_a)$ for all $u \in \tilde{\mathcal{U}}_a$, $f_a \in \tilde{\mathcal{F}}_a$, and $a \in \{2, \dots, \bar{a}\}$, where $\tilde{\mathcal{F}}_a$ is a finite subset of $\mathcal{F}_a$ and $\tilde{\mathcal{U}}_a$ is a finite subset of $\mathcal{U}_a$ chosen by the econometrician.[14] Then any $\theta \in \Theta_{IC}$ satisfies

condition. We require neither condition.

[13] To be precise, the bounds defined by her equations (24) and (25) are feasible, but equation (26) is not. Since we allow for component externalities, the analog of inequality (26) replaces A's neighborhood, "B_A," in Sheng's notation, with the component of A, which we denote by $C_A(G)$. Estimating the moment $\frac{1}{n}\sum_A \mathbf{P}\left((G_A, C_A(G)) = (g_A, C_A(g))\right)$ is impractical since most sets of nodes A will have widely varying components. Thus, her bounds (24) and (25) are the only relevant ones in a large-market setting.

[14] Reducing the infinite set of moments to a finite set in such a fashion clearly entails an additional

$\limsup_{n\to\infty} \mathbf{E}[\bar{m}_a^n(\theta; u, f_a)] \leqslant 0$ for any $u \in \mathcal{U}_a$. These moment inequalities can be used to estimate an identified set by computing, for example,

$$\hat{\Theta}_{IC} = \left\{\theta \in \Theta : ||\bar{m}_a^n(\theta)||_+^2 \leqslant \frac{\log n}{n}\right\},$$

where $\bar{m}_a^n = \left(\bar{m}_a^n(\theta; u, f_a); u \in \tilde{\mathcal{U}}_a, f_a \in \tilde{\mathcal{F}}_a\right)$ and $||x||_+ = ||\max\{x, 0\}||_E$, with $||\cdot||_E$ denoting the Euclidean norm (Chernozhukov et al., 2007). Several optimization algorithms can be used to compute this set. For example, Ciliberto and Tamer (2009) compute use simulated annealing, while Beresteanu et al. (2011) suggest differential evolution (Storn and Price, 1997), which we also use in our empirical application. A number of procedures exist for constructing confidence intervals (e.g. Andrews and Shi, 2013; Bugni et al., 2014; Chernozhukov et al., 2007; Pakes et al., 2011; Romano and Shaikh, 2008; Wan, 2013).

Remark 4 (Variance Estimator)**.** In order to construct confidence intervals, we require a consistent estimate of the variance of $\bar{m}_a^n(\theta; u, f_a)$. We use the HAC estimator due to Jenish (2013). Let $K : \mathbb{R}^{d_\rho} \to [-1, 1]$ be symmetric and continuous at zero and satisfy $K(0) = 1$, $K(x) = 0$ for x such that $|x| > 1$, and $\int |K(x)| dx < \infty$. Then the HAC estimator $\mathcal{V}_a^n(\theta; u, f_a)$ is defined as

$$\frac{1}{n d_a^2 M_n} \sum_{i=1}^n \sum_{\substack{j \in \mathcal{N}_n: \\ ||\delta_{ij}|| \leqslant \tau_n}} K\left(\frac{||\delta_{ij}||}{\tau_n}\right) \left(m_{i,a}^n(\theta; u, f_a) - \bar{m}_a^n(\theta; u, f_a)\right) \left(m_{j,a}^n(\theta; u, f_a) - \bar{m}_a^n(\theta; u, f_a)\right),$$

where the bandwidth τ_n satisfies $\tau_n^{d_\rho} = O(n^{1/3})$. Proposition B.3 proves the consistency of the estimator.

4 Conclusion

This paper studies inference for models of network formation with externalities when the econometrician only observes a small number of networks. Inference is a difficult problem because network externalities can generate dependence among links. We

loss of sharpness. Andrews and Shi (2013) provide an approach that can leverage all instruments in $\mathcal{F}_a$. Beresteanu et al. (2011) extend Andrews and Shi (2013) to leverage all $u \in \mathcal{U}_a$. It is possible to apply these approaches to our setting, but the cost is that the procedures are computationally intensive because they require integrating or maximizing over all u and f_a.

demonstrate that interpretable conditions exist under which such dependence is limited under weak restrictions on externalities. A key assumption is that individuals are homophilous, which limits the probability that dissimilar nodes are connected. Under the assumption that unconnected groups of nodes form links independently, this ensures that links involving highly dissimilar nodes are less correlated, so networks with sufficient diversity have enough independent components. These conditions establish a LLN and CLT for network statistics that can be used to estimate subnetwork moments.

We use these moments to construct new moment inequalities that can be used for inference on θ_0. These moment inequalities are more informative and computationally feasible under weaker restrictions on network externalities than feasible alternatives. The moments are also be useful for estimating games of complete information with moderately many players. Existing moment inequalities in the literature on empirical games require computing the set of equilibria for every candidate parameter, a procedure that suffers from a curse of dimensionality in the number players. Our procedure avoids having to compute equilibria and therefore is feasible when the number of players is large.

A Appendix: Isolated Nodes

Networks often contain a non-negligible share of isolated nodes, that is, nodes with no links. For example, in our application, 24 percent of nodes are isolated. Models following the standard discrete choice setup in which ζ_{ij} is i.i.d., has full support, and is additively separable cannot rationalize this stylized fact, since as $n \to \infty$, with probability approaching one, ζ_{ij} will be arbitrarily large for some j, and hence, node i will be linked to some node j. Here we propose a model with a separable error structure that can rationalize isolates on directed graphs. Let $X_i^u = (\alpha_i, \varrho_i)$, where ϱ_i is Bernoulli distributed, and $X^o = (X_i^o; i \in \mathcal{N}_n)$. Consider the model

$$V_{ij}^n(G, W; \theta) = \tilde{V}_{ij}^n(G, X^o; \theta) + \alpha_i + \varrho_i \zeta_{ij}.$$

We might interpret ζ_{ij} as idiosyncratic meeting shocks between pairs of individuals and ϱ_i as a latent preference for solitude. For simplicity, assume that $\tilde{V}_{ij}^n(\cdot)$ is uniformly bounded above by $\bar{V}$. Then node i is isolated if $\varrho_i = 0$ and $\alpha_i < -\bar{V}$, an event whose probability is uniformly bounded away from zero. Hence, this error structure ensures that the model is consistent with the existence of many isolates, while still preserving the common structure of additively separable unobservables. The same structure can rationalize isolates on undirected graphs for models with "non-transferable utility," where

$$G_{ij} = \mathbf{1}\big\{\tilde{V}^n_{ij}(G, X^o; \theta) + \alpha_i + \varrho_i\zeta_{ij} \geqslant 0\big\}\mathbf{1}\big\{\tilde{V}^n_{ji}(G, X^o; \theta) + \alpha_j + \varrho_j\zeta_{ji} \geqslant 0\big\}.$$

B Appendix: Additional Results

In this section, we consider the case in which G is undirected. The proofs can be easily extended to the directed case.

B.1 Scaling Constants

The first lemma shows that for $M_n = 1$,

$$\frac{1}{nM_n}\sum_{i=1}^n \left(m^n_{i,a}(\theta; u, f_a) - \mathbf{E}[m^n_{i,a}(\theta; u, f_a)]\right) \xrightarrow{p} 0.$$

Define $A{+}i = A \cup \{i\}$.

Proposition B.1. *For any $g_a \in \mathbf{G}^{\leftrightarrow}_{u,a}$, $u \in \mathcal{U}_a$, $f_a \in \mathcal{F}_a$, and $a > 1$, under Assumptions 1, 2, 4, and 5 with $\tau(r) = Ce^{-\varphi r}$ for some $C, \varphi > 0$,*

$$\sup_n \max_{i\in\mathcal{N}_n}\left|\sum_{\substack{A\subseteq\mathcal{N}_n: i\notin A\\ |A|=a-1}} \mathbf{E}\left[\mathbf{1}\left\{G_{A+i} \in g_a\right\} f_a(X^o_{A+i})\right]\right| < \infty \quad \textit{and} \tag{11}$$

$$\sup_n \max_{i\in\mathcal{N}_n}\left|\sum_{\substack{A\subseteq\mathcal{N}_n: i\notin A\\ |A|=a-1}} \mathbf{E}_\theta\left[\max_{g_{-(A+i)}} \max_{g_{A+i}\in\mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u'\mathbb{I}_a(g_{A+i}) f_a(X^o_{A+i})\right]\right| < \infty. \tag{12}$$

Proof. First consider (11) for $a = 3$, so $\mathcal{G}^{\leftrightarrow}_{u,a} = \{\text{intransitive triads, transitive triads}\}$. Since $\mathcal{F}_a$ is uniformly bounded by one and non-negative by assumption, (11) is bounded by

$$\sup_n \max_{i\in\mathcal{N}_n} \sum_{\substack{A\subseteq\mathcal{N}_n: i\notin A\\ |A|=a-1}} \mathbf{E}\left[\mathbf{1}\left\{G_{A+i} \in g_a\right\}\right].$$

Using (1),

$$\mathbf{E}\left[\mathbf{1}\left\{G_{A+i} = g_{A+i}\right\}\right] \leqslant \mathbf{E}\left[\prod_{s,t\in A} \mathbf{1}\left\{\max_{g_{-(A+i)}} (-1)^{1-g_{st}} V^n_{st}\left((g_{A+i}, g_{-(A+i)}), W_{A+i}; \theta_0\right) \geqslant 0\right\}\right],$$

where g_{st} is the st-th component of g_{A+i}. Define for all $i, j, k \in \mathcal{N}_n$ the events

$$\begin{aligned}
\mathcal{I}_{ijk} = \Big\{ & \max_{g_{-\{i,j,k\}}} V_{ij}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}, W;\theta\big) \geqslant 0, \\
& \max_{g_{-\{i,j,k\}}} V_{jk}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}), W;\theta\big) \geqslant 0, \min_{g_{-\{i,j,k\}}} V_{ik}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}), W;\theta\big) < 0\Big\}, \\
\mathcal{T}_{ijk} = \Big\{ & \max_{g_{-\{i,j,k\}}} V_{ij}\big((g^{\mathcal{T}}_{ijk}, g_{-\{i,j,k\}}), W;\theta\big) \geqslant 0, \\
& \max_{g_{-\{i,j,k\}}} V_{jk}\big((g^{\mathcal{T}}_{ijk}, g_{-\{i,j,k\}}), W;\theta\big) \geqslant 0, \min_{g_{-\{i,j,k\}}} V_{ik}\big((g^{\mathcal{T}}_{ijk}, g_{-\{i,j,k\}}), W;\theta\big) \geqslant 0\Big\},
\end{aligned} \tag{13}$$

where $g^{\mathcal{I}}_{ijk}$ is the subnetwork on $\{i, j, k\}$ such that $g_{ij} = g_{jk} = 1 - g_{ik} = 1$, and $g^{\mathcal{T}}_{ijk}$ is the transitive triad on $\{i, j, k\}$. Consider the case in which g_3 is the class of intransitive triads (the argument for transitive triads is similar). Then

$$\begin{aligned}
\sum_{\substack{A\subseteq\mathcal{N}_n: i\notin A \\ |A|=a-1}} \sum_{g_{A+i}\in g_a} \mathbf{E}\left[\prod_{s,t\in A} \mathbf{1}\left\{\max_{g_{-(A+i)}} (-1)^{1-g_{st}} V^n_{st}\left((g_{A+i}, g_{-(A+i)}), W_{A+i};\theta_0\right) \geqslant 0\right\}\right] \\
\leqslant \sum_{\{j,k\}\subseteq\mathcal{N}_n} \left[\mathbf{P}\left(\mathcal{I}_{ijk}\right) + \mathbf{P}\left(\mathcal{I}_{jki}\right) + \mathbf{P}\left(\mathcal{I}_{kij}\right)\right].
\end{aligned} \tag{14}$$

To simplify notation, for any $i, j, k \in \mathcal{N}_n$, define the events

$$\begin{aligned}
\{ij\} &= \{Z_{ij} \in \mathcal{C}(||\delta_{ij}||), Z_{jk} \notin \mathcal{C}(||\delta_{jk}||), Z_{ik} \notin \mathcal{C}(||\delta_{ik}||)\}, \\
\{ij, jk\} &= \{Z_{ij} \in \mathcal{C}(||\delta_{ij}||), Z_{jk} \in \mathcal{C}(||\delta_{jk}||), Z_{ik} \notin \mathcal{C}(||\delta_{ik}||)\}, \text{and} \\
\{ij, jk, ik\} &= \{Z_{ij} \in \mathcal{C}(||\delta_{ij}||), Z_{jk} \in \mathcal{C}(||\delta_{jk}||), Z_{ik} \in \mathcal{C}(||\delta_{ik}||)\}.
\end{aligned} \tag{15}$$

We will only derive the bound for the first term on the right-hand side of (14), as the argument for the others are similar. By the law of total probability, this term equals

$$\begin{aligned}
\sum_{\{j,k\}\subseteq\mathcal{N}_n} \Big(& \mathbf{P}\left(\mathcal{I}_{ijk} \,|\, ij, jk, ik\right) \mathbf{P}(ij, jk, ik) + \mathbf{P}\left(\mathcal{I}_{ijk} \,|\, ij\right) \mathbf{P}(ij) \\
& + \mathbf{P}\left(\mathcal{I}_{ijk} \,|\, ik\right) \mathbf{P}(ik) + \cdots + \text{ etc.}\Big)
\end{aligned} \tag{16}$$

We will only bound the first three terms of this sum, as the argument for the others are similar. The strategy for the bound is as follows. We can write

$$\begin{aligned}\mathbf{P}\left(\mathcal{I}_{ijk}\,\middle|\,ij\right) = \mathbf{E}\Bigg[&\mathbf{1}\big\{\max_{g_{-\{i,j,k\}}} V_{ij}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}, W;\theta\big) \geqslant 0\big\} \\ &\times \mathbf{1}\big\{\max_{g_{-\{i,j,k\}}} V_{jk}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}), W;\theta\big) \geqslant 0\big\} \\ &\mathbf{1}\big\{\min_{g_{-\{i,j,k\}}} V_{ik}\big((g^{\mathcal{I}}_{ijk}, G_{-\{i,j,k\}}), W;\theta\big) < 0\big\}\,\big|\,ij\Bigg].\end{aligned}$$

When conditioning on only ij, replace the indicators containing $V_{jk} \geqslant 0$ and $V_{ik} \geqslant 0$ with their upper bounds, namely one, since these indicators involve pairs (j,k) and (i,k) not equal to (i,j). Likewise, when conditioning on ij, jk, replace the indicator containing $V_{ik} \geqslant 0$ with one. And so on. Replace indicators for $V_{st} < 0$ with one for any $s,t \in \mathcal{N}_n$. Replace $\mathbf{P}(ij)$ with its upper bound $\mathbf{P}\big(Z_{jk} \notin \mathcal{C}(||\delta_{jk}||), Z_{ik} \notin \mathcal{C}(||\delta_{ik}||)\big)$, dropping the events in which attributes lie in their constraint sets, in this case, $\{Z_{ij} \in \mathcal{C}(||\delta_{ij}||)\}$. Likewise with the other events. This strategy leads to a useful bound because g_{A+i} is connected.

Using Assumption 2, the sum of the first three terms of (16) is no greater than

$$\begin{aligned}&\sum_j \mathbf{1}\left\{\sup_{\mathcal{E}\in E}\sup_{z\in\mathcal{C}(||\delta_{ij}||)} V_{ij}(\delta_{ij},\mathcal{E},z;\theta_0) \geqslant 0\right\}\sum_k \mathbf{1}\left\{\sup_{\mathcal{E}\in E}\sup_{z\in\mathcal{C}(||\delta_{jk}||)} V_{jk}(\delta_{jk},\mathcal{E},z;\theta_0) \geqslant 0\right\} \\ &\quad+\sum_j \mathbf{1}\left\{\sup_{\mathcal{E}\in E}\sup_{z\in\mathcal{C}(||\delta_{ij}||)} V_{ij}(\delta_{ij},\mathcal{E},z;\theta_0) \geqslant 0\right\}\sum_k \mathbf{P}\big(Z_{jk}\notin\mathcal{C}(||\delta_{jk}||), Z_{ik}\notin\mathcal{C}(||\delta_{ik}||)\big) \\ &\qquad+\sum_j\sum_k \mathbf{P}\big(Z_{ij}\notin\mathcal{C}(||\delta_{ij}||), Z_{jk}\notin\mathcal{C}(||\delta_{jk}||)\big).\end{aligned} \tag{17}$$

We will show that these terms are uniformly bounded.

Consider the first summand of (17). For all $j \in \mathcal{N}_n$, define $S_{j,t} = \{k \in \mathcal{N}_n : ||\delta_{jk}|| \in [t, t+1)\}$. Then

$$\begin{aligned}&\sum_k \mathbf{1}\left\{\sup_{\mathcal{E}\in E}\sup_{z\in\mathcal{C}(||\delta_{jk}||)} V_{jk}(\delta_{jk},\mathcal{E},z;\theta_0) \geqslant 0\right\} \\ &\qquad\leqslant \sum_{t=0}^{\infty} |S_{j,t}|\mathbf{1}\left\{\max_{j,k\in\mathbb{N}}\sup_{\mathcal{E}\in E}\sup_{z\in\mathcal{C}(||t||)} V_{jk}(t,\mathcal{E},z;\theta_0) \geqslant 0\right\}.\end{aligned} \tag{18}$$

By Assumption 5 and Lemma A.1 of Jenish and Prucha (2009), $|S_{j,t}|$ is uniformly $O(t^{d_\rho - 1})$. By Assumption 4(b), the indicator function is zero for t sufficiently large. Hence, the right-hand side converges.

This argument implies that the first term of (17) is uniformly bounded. For the second term, notice

$$\sum_k \mathbf{P}\big(Z_{jk} \notin \mathcal{C}(||\delta_{jk}||), Z_{ik} \notin \mathcal{C}(||\delta_{ik}||)\big) \leqslant \sum_k \mathbf{P}\big(Z_{jk} \notin \mathcal{C}(||\delta_{jk}||)\big)$$
$$\leqslant \sum_{t=1}^{\infty} |S_{j,t}|\big(\mathbf{P}(X_j \notin \mathcal{C}_x(t)) + \mathbf{P}(X_k \notin \mathcal{C}_x(t)) + \mathbf{P}(\zeta_{jk} \notin \mathcal{C}_\zeta(t))\big). \quad (19)$$

By Assumption 4(a), the probabilities are all $O(t^{-5d_\rho-\varphi})$. This and the argument for (18) imply that the second term of (17) is uniformly bounded, as desired.

Lastly the third term of (17) can be bounded using the Cauchy-Schwarz inequality and arguments for the other terms. (Note that for large values of a, this part of the argument will need to make use of the fact that attributes have exponential tails, but for any d_ρ, exponential tails are unnecessary for $a \leqslant 3$.) This completes the proof of (11).

Turning to (12), as with (11), we drop $f_a(X^o_{A+i})$ from the expression without loss of generality, and consider the case $a = 3$. For $a = 3$, $\mathbf{G}^{\leftrightarrow}_{u,a}$ contains two elements, the equivalence classes of intransitive and transitive triads. Let $u = (u_1, u_2)$. Then

$$\Bigg| \sum_{\substack{A \subseteq \mathcal{N}_n : i \notin A \\ |A| = a-1}} \mathbf{E}_\theta \left[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_A(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) \right] \Bigg|$$
$$\leqslant \Bigg| \sum_{j,k} \mathbf{E}_\theta \Bigg[\Bigg(\max\{u_1, u_2\} \mathbf{1}\{\mathcal{I}_{ijk} \cap \mathcal{T}_{ijk}\}$$
$$+ u_1 \mathbf{1}\{\mathcal{I}_{ijk} \cap \mathcal{T}^c_{ijk}\} + u_2 \mathbf{1}\{\mathcal{I}^c_{ijk} \cap \mathcal{T}_{ijk}\} \Bigg) \Bigg] \Bigg|$$

This is bounded by

$$\sum_{j,k} \mathbf{E}_\theta \left[3 \,|\max\{u_1, u_2\}| \mathbf{1}\{\mathcal{I}_{ijk} \cup \mathcal{T}_{ijk}\} \right]$$
$$\leqslant 3|\max\{u_1, u_2\}| \left(\sum_{j,k} \mathbf{E}_\theta \left[\mathbf{1}\{\mathcal{I}_{ijk}\} \right] \right.$$
$$\left. + \sum_{j,k} \mathbf{E}_\theta \left[\mathbf{1}\{\mathcal{T}_{ijk}\} \right] \right). \quad (20)$$

Both of the sums on the right-hand side are uniformly bounded by arguments for (14).

This completes the proof for $a = 3$. The proofs for other values of a can be derived using the same arguments. The intuition is that the expected degree of a node is uniformly bounded in this model (this is implied by the $a = 2$ case), so the average number of connected subnetworks g_{A+i} containing i should be similarly bounded. ■

B.2 Hoeffding Decomposition

Fix a. As previously noted, the second term of (10), averaged over all subsets of nodes A of size a, is a U-statistic whose kernel is order a. The next proposition proves the familiar decomposition of U-statistics into an average of independent terms and an asymptotically negligible remainder term. This requires a new argument relative to the standard case because, in light of Proposition B.1, the U-statistic is scaled by $\frac{1}{n}$, rather than $\binom{n}{a}^{-1}$. For $A = \{i_1, \dots, i_a\}$, and fixed $u \in \mathcal{U}_a$ and $\theta \in \Theta$, define the kernel

$$J_A(X_{i_1}, \dots, X_{i_a}) = J_A(Z_A^o) = \mathbf{E}_\theta \left[\max_{g_{-A}} \max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)} u' \mathbb{I}_a(g_A) f_a(Z_A^o) \,\middle|\, Z_A^o \right].$$

Let $U_n = \sum_{A \subseteq \mathcal{N}_n, |A|=a} J_A(Z_A^o)$, and define its projection

$$\begin{aligned} S_n &= \sum_{i=1}^n \big(\mathbf{E}_\theta[U_n \,|\, X_i^o] - \mathbf{E}_\theta[U_n] \big) + \mathbf{E}_\theta[U_n] \\ &= \frac{1}{a} \sum_{i=1}^n \sum_{\substack{A \subseteq \mathcal{N}_n : i \notin A \\ |A| = a-1}} \left(a\, \mathbf{E}_\theta \left[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) f_a(X_{A+i}^o) \,\middle|\, X_i^o \right] \right. \\ &\qquad \left. - (a-1) \mathbf{E}_\theta \left[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) f_a(X_{A+i}^o) \right] \right). \end{aligned}$$

Proposition B.2. *Under the assumptions of Proposition B.1, $\frac{1}{n} U_n = \frac{1}{n} S_n + o_p(n^{-1/2})$.*

Proof. It suffices to show that $\frac{1}{\sqrt{n}}(S_n - U_n) \xrightarrow{L_2} 0$. Below, we prove that $\mathrm{Var}\left(\frac{1}{\sqrt{n}} U_n\right) < \infty$, Supposing for now that this is true, by Hájek's projection lemma (Hájek, 1968),

$$\begin{aligned} \mathbf{E}\left(\frac{1}{\sqrt{n}} S_n - \frac{1}{\sqrt{n}} U_n \right)^2 &= \frac{1}{n} \left(\mathrm{Var}(S_n) - \mathrm{Var}(U_n) \right) \\ &= \frac{1}{n} \mathrm{Var}(U_n) \left(\frac{\mathrm{Var}(S_n)}{\mathrm{Var}(U_n)} - 1 \right). \end{aligned}$$

Since the collection of kernels $\{J_A\}_{A \subseteq \mathcal{N}_n}$ is uniformly bounded, by the arguments in Theorem 1, section 3.7.2 of Lee (1990), $\binom{n}{a}^{-2} \mathrm{Var}(S_n) / \binom{n}{a}^{-2} \mathrm{Var}(U_n) \to 1$. Thus, it remains to show that $\mathrm{Var}(U_n) = O(n)$, or equivalently, $\mathrm{Var}\left(\frac{1}{\sqrt{n}} U_n \right) < \infty$

As in Proposition B.1, we will only consider the case in which the kernel has order $a = 3$. Similar arguments apply for kernels of other orders. Notice

$$\mathrm{Var}(U_n) = \sum_{c=1}^{a} \sum_{|A\cap B|=c} \mathrm{Cov}\left(J_A(Z_A^o), J_B(X_B)\right).$$

Then it suffices to show that for any $c \geqslant 1$, $\sum_{|A\cap B|=c} \mathrm{Cov}\left(J_A(Z_A^o), J_B(Z_B^o)\right) = O(n)$. The strategy is the same as the proof of Proposition B.1. First we examine the case of $c = 1$. For ease of notation, define $J_{ijk} = J_{\{i,j,k\}}(X^o_{\{i,j,k\}})$. Then adopting the notation in equation (16) of Proposition B.1, by the law of total probability,

$$\begin{aligned}
\sum_{|A\cap B|=1} \mathrm{Cov}\left(J_A(Z_A^o), J_B(X_B)\right) &= \frac{1}{3!2!} \sum_{i=1}^{n} \sum_{j\neq k\neq i} \sum_{l\neq m\neq i} \big(\mathbf{E}[J_{ijk}J_{ilm}] - \mathbf{E}[J_{ijk}]\mathbf{E}[J_{ilm}]\big) \\
&= \frac{1}{12} \sum_{i=1}^{n} \sum_{j\neq k\neq i} \sum_{l\neq m\neq i} \big(\mathbf{E}[J_{ijk}J_{ilm} \,|\, il]\mathbf{P}(il) + \mathbf{E}[J_{ijk}J_{ilm} \,|\, ij, il]\mathbf{P}(ij, il) \\
&\quad + \mathbf{E}[J_{ijk}J_{ilm} \,|\, ik, il]\mathbf{P}(ik, il) + \mathbf{E}[J_{ijk}J_{ilm} \,|\, ij, ik, il]\mathbf{P}(ij, ik, il) + \cdots + \text{etc.}\big) \\
&\quad + \frac{1}{12} \sum_{i=1}^{n} \sum_{j\neq k\neq i} \sum_{l\neq m\neq i} \big(\mathbf{E}[J_{ijk} \,|\, ij]\mathbf{P}(ij) + \mathbf{E}[J_{ijk} \,|\, ij, ik]\mathbf{P}(ij, ik) + \cdots + \text{etc.}\big) \\
&\qquad \times \big(\mathbf{E}[J_{ilm} \,|\, il]\mathbf{P}(il) + \mathbf{E}[J_{ilm} \,|\, il, im]\mathbf{P}(il, im) + \cdots + \text{etc.}\big).
\end{aligned}$$

The summation spanning the last two lines is $O(n)$ by the arguments for (16). To show that the summation immediately preceding it is also $O(n)$, we consider as an example the term

$$\sum_{i=1}^{n} \sum_{j\neq k\neq i} \sum_{l\neq m\neq i} \mathbf{E}[J_{ijk}J_{ilm} \,|\, ik, il]\mathbf{P}(ik, il).$$

(The argument for the other terms follow similarly.) The ith summand of the outer sum is bounded in absolute value by

$$\begin{aligned}
\sum_{j\neq k\neq i} \sum_{l\neq m\neq i} \mathbf{E}\Big[\mathbf{E}_\theta\Big[\mathbf{1}\{\mathcal{I}_{ijk}(g_{-\{i,j,k\}})\} + \mathbf{1}\{\mathcal{T}_{ijk}(g_{-\{i,j,k\}})\} \,\Big|\, X^o_{\{i,j,k\}}\Big] \\
\times \mathbf{E}_\theta\Big[\mathbf{1}\{\mathcal{I}_{ilm}(g_{-\{i,j,k\}})\} + \mathbf{1}\{\mathcal{T}_{ilm}(g_{-\{i,j,k\}})\} \,\Big|\, X^o_{\{i,l,m\}}\Big] \,\Big|\, ik, kl\Big] \\
\times 9\max\{u_1, u_2\}^2\, \mathbf{P}(ik, il),
\end{aligned}$$

using the argument in (20). Using the strategy for bounding (16), we can bound this by

$9 \max\{u_1, u_2\}^2$ times

$$\sum_{j \neq k} 2 \cdot \mathbf{1}\left\{\sup_{\mathcal{E} \in E} \sup_{z \in \mathcal{C}(||\delta_{ik}||)} V_{ik}(\delta_{ik}, \mathcal{E}, z; \theta_0) \geqslant 0\right\} \mathbf{P}\left(\zeta_{ik} \notin \mathcal{C}_\zeta(||\delta_{ik}||), \zeta_{jk} \notin \mathcal{C}_\zeta(||\delta_{jk}||)\right)$$
$$\times \sum_{l \neq m} 2 \cdot \mathbf{1}\left\{\sup_{\mathcal{E} \in E} \sup_{z \in \mathcal{C}(||\delta_{il}||)} V_{il}(\delta_{il}, \mathcal{E}, z; \theta_0) \geqslant 0\right\} \mathbf{P}\left(\zeta_{il} \notin \mathcal{C}_\zeta(||\delta_{il}||), \zeta_{lm} \notin \mathcal{C}_\zeta(||\delta_{lm}||)\right)$$

This is uniformly bounded by arguments for (16).

Next, we consider the case $c = 2$:

$$\sum_{|A \cap B| = 2} \mathrm{Cov}\left(J_A(Z_A^o), J_B(X_B)\right) = \frac{1}{3!} \sum_{i=1}^n \sum_{j \neq k \neq i} \sum_{l \neq i,j} \big(\mathbf{E}[J_{ijk} J_{ilm}] - \mathbf{E}[J_{ijk}]\mathbf{E}[J_{ilm}]\big)$$
$$= \frac{1}{6} \sum_{i=1}^n \sum_{j \neq k \neq i} \sum_{l \neq i,j} \big(\mathbf{E}[J_{ijk} J_{ijl} \,|\, il]\mathbf{P}(il) + \mathbf{E}[J_{ijk} J_{ilm} \,|\, ij, il]\mathbf{P}(ij, il)$$
$$+ \mathbf{E}[J_{ijk} J_{ijl} \,|\, ik, il]\mathbf{P}(ik, il) + \mathbf{E}[J_{ijk} J_{ijl} \,|\, ij, ik, il]\mathbf{P}(ij, ik, il) + \cdots + \text{etc.}\big)$$
$$+ \frac{1}{6} \sum_{i=1}^n \sum_{j \neq k \neq i} \sum_{l \neq i,j} \big(\mathbf{E}[J_{ijk} \,|\, ij]\mathbf{P}(ij) + \mathbf{E}[J_{ijk} \,|\, ij, ik]\mathbf{P}(ij, ik) + \cdots + \text{etc.}\big)$$
$$\times \big(\mathbf{E}[J_{ijl} \,|\, ij]\mathbf{P}(ij) + \mathbf{E}[J_{ijl} \,|\, ij, il]\mathbf{P}(ij, il) + \cdots + \text{etc.}\big).$$

The summation spanning the last two lines is $O(n)$ by the arguments for (16). To show that the summation immediately preceding it is also $O(n)$, we consider as an example the term

$$\sum_{i=1}^n \sum_{j \neq k \neq i} \sum_{l \neq i,j} \mathbf{E}[J_{ijk} J_{ijl} \,|\, jk, ik]\mathbf{P}(jk, ik).$$

(The argument for the other terms follow similarly.) Following the same line of reasoning as the $c = 1$ case, the ith summand of the outer sum is bounded in absolute value by

$$9 \max\{u_1, u_2\}^2 \sum_i \sum_{j \neq k} \mathbf{1}\left\{\sup_{\mathcal{E} \in E} \sup_{z \in \mathcal{C}(||\delta_{jk}||)} V_{jk}(\delta_{jk}, \mathcal{E}, z; \theta) \geqslant 0\right\} \sum_l \mathbf{P}(\zeta_{ij} \notin \mathcal{C}_\zeta(||\delta_{ij}||), \zeta_{il} \notin \mathcal{C}_\zeta(||\delta_{il}||).$$

This is uniformly $O(n)$ by arguments for (17). The case $c = 3$ can be shown with similar arguments. ■

B.3 Variance Estimator

Proposition B.3. *Under the assumptions of Lemma B.1, for any $a \in \mathbb{N}$, $u \in \mathcal{U}_a$, $f_a \in \mathcal{F}_a$, and $\theta \in \Theta$,*

$$\mathcal{V}_a^n(\theta; u, f_a) - Var\big(\sqrt{n}\bar{m}_a^n(\theta; u, f_a)\big) \xrightarrow{p} 0.$$

Proof. Define

$$\begin{aligned} w_{i,a}^n(\theta; u, f_a) = \sum_{\substack{A \subseteq \mathcal{N}_n : i \notin A \\ |A| = a-1}} \Bigg(& u' \mathbb{I}_a(G_{A+i}) f_a(X_{A+i})^o \\ & - a\mathbf{E}_\theta \left[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) f_a(X_{A+i}^o) \,\middle|\, X_i^o \right] \\ & + (a-1)\mathbf{E}_\theta \left[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) f_a(X_{A+i}^o) \right] \Bigg) \end{aligned}$$

and $\bar{w}_a^n(\theta; u, f_a) = \frac{1}{an} \sum_{i=1}^n w_{i,a}^n(\theta; u, f_a)$. By Proposition B.2,

$$\sqrt{n}\bar{m}_a^n(\theta; u, f_a) = \sqrt{n}\bar{w}_a^n(\theta; u, f_a) + o_p(1).$$

Define $\mathcal{W}_a^n(\theta; u, f_a)$ as

$$\frac{1}{a^2 n} \sum_{i=1}^n \sum_{\substack{j \in \mathcal{N}_n : \\ ||\delta_{ij}|| \leq \tau_n}} K\left(\frac{||\delta_{ij}||}{\tau_n}\right) \left(w_{i,a}^n(\theta; u, f_a) - \bar{w}_a^n(\theta; u, f_a)\right) \left(w_{j,a}^n(\theta; u, f_a) - \bar{w}_a^n(\theta; u, f_a)\right),$$

By Jenish (2013) Theorem 4, $\mathcal{W}_a^n(\theta; u, f_a) - \mathrm{Var}\big(\sqrt{n}\bar{w}_a^n(\theta; u, f_a)\big) \xrightarrow{p} 0$. Since $\sqrt{n}\bar{w}_a^n(\theta; u, f_a) - \sqrt{n}\bar{m}_a^n(\theta; u, f_a) \xrightarrow{L_2} 0$, as shown in the proof of Proposition B.2, it follows from the (reverse) triangle inequality that $\mathrm{Var}\big(\sqrt{n}\bar{w}_a^n(\theta; u, f_a)\big) - \mathrm{Var}\big(\sqrt{n}\bar{m}_a^n(\theta; u, f_a)\big) \to 0$. Therefore, $\mathcal{W}_a^n(\theta; u, f_a)$ is consistent for the variance of $\sqrt{n}\bar{m}_a^n(\theta; u, f_a)$. This estimator is not feasible, since it contains an integral over the distribution of elements of X, which is not known. Thus, we aim to show that $\mathcal{W}_a^n(\theta; u, f_a) - \mathcal{V}_a^n(\theta; u, f_a) \xrightarrow{p} 0$, which will establish the proof.

To simplify notation, we fix a and make several definitions:

- $w_i = w_{i,a}^n(\theta; u, f_a)$, $\bar{w} = \frac{1}{an} \sum_{i=1}^n w_i$,
- $m_i = m_{i,a}^n(\theta; u, f_a)$, $\bar{m} = \frac{1}{an} \sum_{i=1}^n m_i$,
- $K_{ij} = K\left(\frac{||\delta_{ij}||}{\tau_n}\right)$,
- $D_i = \sum_{\substack{A \subseteq \mathcal{N}_n : i \notin A \\ |A| = a-1}} u' \mathbb{I}_a(G_{A+i}) f_a(X_{A+i}^o)$,

- $H_i = -\sum_{\substack{A \subseteq \mathcal{N}_n : i \notin A \\ |A| = a-1}} \Bigg(a \Big[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) f_a(X^o_{A+i}) \,\big|\, X^o_i \Big]$
  $-(a-1)\mathbf{E}_\theta \Big[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) f_a(X^o_{A+i}) \Big] \Bigg)$, and
- $U_i = -\sum_{\substack{A \subseteq \mathcal{N}_n : i \notin A \\ |A| = a-1}} \Bigg(\mathbf{E}_\theta \Big[\max_{g_{-(A+i)}} \max_{g_{A+i} \in \mathcal{G}_{A+i}(g_{-(A+i)}, W_{A+i}, \theta)} u' \mathbb{I}_a(g_{A+i}) f_a(X^o_{A+i}) \,\big|\, X^o_{A+i} \Big] \Bigg)$.

So $w_i = D_i + H_i$, and $m_i = D_i + U_i$, and

$$\mathcal{W}^n_a(u_a, f_a, \theta) - \mathcal{V}^n_a(u_a, f_a, \theta) = \underbrace{\frac{1}{a^2 n} \sum K_{ij}(w_i w_j - m_i m_j)}_{\mathsf{A}}$$
$$+ \underbrace{\frac{1}{a^2 n} \sum K_{ij}(m_i \bar{m} - w_i \bar{w})}_{\mathsf{B}} + \underbrace{\frac{1}{a^2 n} \sum K_{ij}(m_j \bar{m} - w_j \bar{w})}_{\mathsf{C}} + \underbrace{\frac{1}{a^2 n} \sum K_{ij}(\bar{w}^2 - \bar{m}^2)}_{\mathsf{D}},$$

where the sum is over all i and j such that $||\delta_{ij}|| \leqslant \tau_n$. We show that each of these components is $o_p(1)$. First,

$$\begin{aligned}
\mathsf{B} &= \frac{1}{a^2 n} \sum_i \sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} \big(m_i \bar{m} - w_i (\bar{m} + \bar{w} - \bar{m}) \big) \\
&= \frac{1}{a^2 n} \sum_i (m_i - w_i) \sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} \bar{m} - \frac{1}{a^2 n} \sum w_i \sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} (\bar{w} - \bar{m}).
\end{aligned}$$

Since K_{ij} is bounded above by one, and $\tau_n^{d_\rho} = O(n^{1/3})$ by construction, by Lemma A.1 of Jenish and Prucha (2009), $\sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} = O(n^{1/3})$. By Proposition B.2, $\bar{w} - \bar{m} = \frac{1}{n} \sum_i (m_i - w_i) = o_p(n^{-1/2})$. Hence, $\mathsf{B} = o_p(1)$. The argument for C is similar. Next,

$$\begin{aligned}
\mathsf{D} &= \frac{1}{a^2 n} \sum_i \sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} \left((\bar{m} + \bar{w} - \bar{m})^2 - \bar{m}^2 \right) \\
&= \frac{2}{a^2 n} \sum_i \sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} \bar{m} (\bar{w} - \bar{m}) + \frac{1}{a^2 n} \sum_i \sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} (\bar{w} - \bar{m})^2.
\end{aligned}$$

The terms in the last line are $o_p(1)$ by the argument for B. Finally,

$$\begin{aligned}
\mathsf{A} &= \frac{1}{a^2 n} \sum K_{ij} \big((D_i + H_i)(D_j + H_j) - (D_i + U_i)(D_j + U_j) \big) \\
&= \underbrace{\frac{1}{a^2 n} \sum K_{ij} D_i (H_j - U_j)}_{\mathsf{A}_1} + \underbrace{\frac{1}{a^2 n} \sum K_{ij} D_j (H_i - U_i)}_{\mathsf{A}_2} + \underbrace{\frac{1}{a^2 n} \sum K_{ij} (H_i H_j - U_i U_j)}_{\mathsf{A}_3}.
\end{aligned}$$

Term $\mathsf{A}_2 = \frac{1}{a^2 n} \sum_i (H_i - U_i) \sum_{j : ||\delta_{ij}|| \leqslant \tau_n} K_{ij} D_j = o_p(1)$ by arguments for B, noting that D_j

is uniformly bounded above by Proposition B.1. Next,

$$\begin{aligned}\mathsf{A}_3 &= \frac{1}{a^2 n}\sum K_{ij}\big(H_iH_j - (H_i+U_i-H_i)(H_j+U_j-H_j)\big)\\ &= \underbrace{-\frac{1}{a^2 n}\sum_i H_i \sum_{j:||\delta_{ij}||\leqslant\tau_n} K_{ij}(U_j-H_j)}_{\mathsf{B}_1} \underbrace{-\frac{1}{a^2 n}\sum_i (U_i-H_i)\sum_{j:||\delta_{ij}||\leqslant\tau_n} K_{ij}H_j}_{\mathsf{B}_2}\\ &\qquad \underbrace{-\frac{1}{a^2 n}\sum_i (U_i-H_i)\sum_{j:||\delta_{ij}||\leqslant\tau_n} K_{ij}(U_j-H_j)}_{\mathsf{B}_3}.\end{aligned}$$

Terms B_2 and B_3 are $o_p(1)$ by arguments for B, noting that U_j and H_j are uniformly bounded above. Notice

$$\mathsf{B}_1 = -\frac{1}{a^2 n}\sum_j (U_j - H_j)\sum_i K_{ij}H_i \mathbf{1}\{||\delta_{ij}|| \leqslant \tau_n\}.$$

By arguments for B, first average is $o_p(n^{-1/2})$ (Proposition B.2), while the last sum is $O_p(n^{1/3})$. Hence, A_2 and A_3 are $o_p(1)$. Similar arguments demonstrate that $\mathsf{A}_1 = o_p(1)$, noting that D_i is uniformly bounded above. This completes the proof. ■

B.4 Verifying Regularity Conditions

We verify Assumptions A.1-A.4 of Bugni et al. (2014). Define the empirical process

$$v_a^n(\theta; u, f_a) = \sigma_a^n(\theta; u, f_a)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(m_{i,a}^n(\theta; u, f_a) - \mathbf{E}\left[m_{i,a}^n(\theta; u, f_a)\right]\right),$$

where $\sigma_a^n(\theta; u, f_a)^2 = \mathrm{Var}\left(\sqrt{n}\bar{m}_a^n(\theta; u, f_a)\right)$. Further, define the correlation matrix

$$\Omega_{a,a'}^n(\theta,\theta'; u, u', f_a, f_{a'}) = \mathbf{E}\left[\left(\frac{\bar{m}_a^n(\theta; u, f_a) - \mathbf{E}[\bar{m}_a^n(\theta; u, f_a)]}{\sigma_a^n(\theta; u, f_a)}\right)\left(\frac{\bar{m}_a^n(u', f'_{a'}, \theta') - \mathbf{E}[\bar{m}_{a'}^n(\theta'; u', f_{a'})]}{\sigma_{a'}^n(\theta'; u', f_{a'})}\right)\right].$$

Let $\mathcal{P}$ be the set of distributions of the primitives W, ν such that Θ_{IC} is nonempty. For $F \in \mathcal{P}$, let $\mathbf{P}_F^*$ denote outer probability and $d_F(\theta, \theta')$ the coordinate-wise intrinsic variance semimetric (equation (A-1) of Bugni et al., 2014). Implicitly, the expectations and probabilities in this section are indexed by F.

We maintain the following assumptions.

Assumption 10. $\liminf_{n\to\infty} \lambda_{min}\left(\sigma_a^n(u, f_a, \theta)^2\right) > 0$, *where* λ_{min} *is the smallest eigenvalue.*

Assumption 11 (Regularity). *Let* $\phi_{A,\theta}(\cdot \mid Z_A^o)$ *be the conditional density of* $\varepsilon_A \mid Z_A^o$. *For any* $i, j \in \mathbb{N}$ *and* $A \subseteq \mathcal{N}_n$, V_{ij} *and* $\phi_{A,\theta}(\cdot \mid Z_A^o)$ *are differentiable in* θ.

The next assumption can likely be relaxed, but this requires a central limit theorem for U-processes with observation-indexed kernel functions. To the best of our knowledge, no such result currently exists, although there has been work establishing central limit theorems for U-statistics with observation-indexed kernels (e.g. De Jong, 1987).

Assumption 12 (Anonymity). *Maintaining Assumptions 2 and 9, for any* $i, j \in \mathcal{N}_n$, *there exist functions* V *and* $\mathcal{E}^n$ *such that*

$$V_{ij}(\delta_{ij}, \mathcal{E}_{ij}^n(G, Z_{ij}), Z_{ij}; \theta) = V(\delta_{ij}, \mathcal{E}^n(G_i, G_j, G_{-i,-j}, Z_{ij}), Z_{ij}; \theta),$$

where $G_{-i,-j}$ *is the subnetwork* G *with all links involving nodes* i *and* j *removed.*

In other words, V_{ij} does not depend directly on the labels i and j. This assumption rules out the possibility that different nodes can have different utility functions, interpreted as different "roles" in the network-formation process. This is satisfied by all of the examples in this paper.

Lemma B.1. *Under Assumptions 2, 9, 10, 11, and 12,*

$$\mathbf{E}_\theta\left[\max_{g_{-A}} \max_{g_A \in \mathcal{G}_{A+i}(g_{-A}, W_A, \theta)} u' \mathbb{I}_a(g_A) f_a(Z_A^o) \,\middle|\, Z_A^o\right] \tag{21}$$

is differentiable in θ *for any* $A \subseteq \mathcal{N}_n$.

Proof. We suppress $f_a(Z_A^o)$ for ease of notation. We show the claim for the case $a = 3$;

the proof is similar for other values. Let $A = \{i, j, k\}$, and define

$$\mathcal{I}_{ijk}(g_{-A}) = \Big\{ V_{ij}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}, W; \theta\big) \geqslant 0,$$
$$V_{jk}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}), W; \theta\big) \geqslant 0, V_{ik}\big((g^{\mathcal{I}}_{ijk}, g_{-\{i,j,k\}}), W; \theta\big) < 0 \Big\},$$
$$\mathcal{T}_{ijk}(g_{-A}) = \Big\{ V_{ij}\big((g^{\mathcal{T}}_{ijk}, g_{-\{i,j,k\}}), W; \theta\big) \geqslant 0,$$
$$V_{jk}\big((g^{\mathcal{T}}_{ijk}, g_{-\{i,j,k\}}), W; \theta\big) \geqslant 0, V_{ik}\big((g^{\mathcal{T}}_{ijk}, g_{-\{i,j,k\}}), W; \theta\big) \geqslant 0 \Big\}.$$

Since $a = 3$, vectors in $\mathcal{U}_a$ are two dimensional, and we label each such u as (u_1, u_2). Expression (21) can be written as

$$\mathbf{E}_\theta \left[\max_{g_{-A}} \Bigg(\max\{u_1, u_2\} \mathbf{1}\left\{ \mathcal{I}_{ijk}(g_{-A}) \cap \mathcal{T}_{ijk}(g_{-A}) \right\} \right.$$
$$\left. + u_1 \mathbf{1}\left\{ \mathcal{I}_{ijk}(g_{-A}) \cap \mathcal{T}_{ijk}(g_{-A})^c \right\} + u_2 \mathbf{1}\left\{ \mathcal{I}_{ijk}(g_{-A})^c \cap \mathcal{T}_{ijk}(g_{-A}) \right\} \Bigg) \,\Bigg|\, Z^o_A \right].$$

To reduce notation in the next expression, we will define $\mathcal{I} = \mathcal{I}_{ijk}(g_{-A})$ and $\mathcal{T} = \mathcal{T}_{ijk}(g_{-A})$. Then the previous equation equals the conditional expectation under θ of

$$\begin{aligned}
&\mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\} \max\{u_1, u_2\} + \\
&\mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\} \max\{u_1, u_2\} + \\
&\mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\} \max\{u_1, u_2\} + \\
&\mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\} \max\{u_1, u_2\} + \\
&\mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\} \max\{u_1, u_2\} + \\
&\mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, u_1 + \\
&\mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\nexists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, u_2.
\end{aligned}$$

Consider the expectation of the first element of the sum in the previous expression (the argument for the other elements is the same):

$$\mathbf{E}_\theta \left[\mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}\, \mathbf{1}\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\} \max\{u_1, u_2\} \,\middle|\, Z^o_A \right]$$
$$= \max\{u_1, u_2\} \int_{\varepsilon_A :\, \mathcal{L}} \phi_{A,\theta}(\varepsilon_A \,|\, Z^o_A)\, d\varepsilon_A,$$

where the event $\mathcal{L}$ equals

$$\{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}\} \cap \{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\} \cap \{\exists\, g_{-A} : \mathcal{I} \cap \mathcal{T}^c\}.$$

By the Leibniz integral rule, the integral is differentiable in θ since the limits of integration and the density are differentiable under Assumption 11. This completes the proof of (a). ■

Assumption 13. $\sup_n \sup_{\theta,\theta'} \sup_{F\in\mathcal{P}} \nabla_{\theta,\theta'} ||\Omega^n_{a,a'}(\theta,\theta'; u, u', f_a, f_{a'})|| < \infty$. *(This derivative is well defined by Lemma B.1.)*

Lemma B.2. *Maintain the assumptions of Lemma B.1. Under Assumption 13, for any* $a, a' > 1$, $u \in \mathcal{U}_a$, $u' \in \mathcal{U}_{a'}$, $f_a \in \tilde{\mathcal{F}}_a$, *and* $f'_{a'} \in \tilde{\mathcal{F}}_{a'}$,

$$\lim_{\delta\downarrow 0} \sup_{||(\theta_1,\theta_1')-(\theta_2,\theta_2')||<\delta} \sup_{F\in\mathcal{P}} ||\Omega^n_{a,a'}(\theta_1,\theta_1'; u, u', f_a, f_{a'}) - \Omega^n_{a,a'}(\theta_2,\theta_2'; u, u', f_a, f_{a'})|| = 0.$$

Proof. This follows from a first-order Taylor expansion of the correlation matrix. ■

Proposition B.4 (Stochastic Equicontinuity)**.** *Under Assumptions 2, 9, 10, 11, and 12,*

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{F\in\mathcal{P}} \mathbf{P}^*_F \left(\sup_{d_F(\theta,\theta')<\delta} ||v_n(\theta) - v_n(\theta')|| > \epsilon \right) = 0.$$

Proof. Recall the definitions of D_i and U_i from the proof of Proposition B.3. To emphasize its dependence on θ, we write $U_i(\theta) \equiv U_i$. Define $m^n_{i,a} \equiv m^n_{i,a}(\theta; u, f_a)$, $\sigma^n_a(\theta) \equiv \sigma^n_a(\theta; u, f_a) = \mathrm{Var}\,(\sqrt{n}\bar{m}^n_a(\theta; u, f_a))$, and $v_n(\theta) \equiv v_n(\theta; u, f_a) = \sigma^n_a(\theta)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(m^n_{i,a}(\theta) - \mathbf{E}[m^n_{i,a}(\theta)] \right)$. Notice

$$\begin{aligned} v_n(\theta) - v_n(\theta') = \sigma^n_a(\theta)^{-1} \left(\sigma^n_a(\theta) v_n(\theta) - \sigma^n_a(\theta') v_n(\theta') \right) \\ + \left(\sigma^n_a(\theta)\sigma^n_a(\theta')^{-1} - 1 \right) \left(\frac{1}{\sqrt{n}} \sum_i (D_i - \mathbf{E}[D_i]) \right) \\ + \left(\sigma^n_a(\theta)\sigma^n_a(\theta')^{-1} - 1 \right) \left(\frac{1}{\sqrt{n}} \sum_i \left(U_i(\theta') - \mathbf{E}[U_i(\theta')] \right) \right). \end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbf{P}_F^*\left(\sup_{d_F(\theta,\theta')<\delta}||v_n(\theta)-v_n(\theta')||>\epsilon\right)\\
&\qquad\leqslant \mathbf{P}_F^*\left(\sup_{d_F(\theta,\theta')<\delta}\left|\left|\sigma_a^n(\theta)^{-1}\left(\sigma_a^n(\theta)v_n(\theta)-\sigma_a^n(\theta')v_n(\theta')\right)\right|\right|>\frac{\epsilon}{3}\right)\\
&\qquad+\mathbf{P}_F^*\left(\sup_{d_F(\theta,\theta')<\delta}\left|\left|\left(\sigma_a^n(\theta)\sigma_a^n(\theta')^{-1}-1\right)\right|\right|\times\left|\left|n^{-1}\mathrm{Var}\left(\sum_i D_i\right)\right|\right|\right.\\
&\qquad\qquad\left.\times\left|\left|n\mathrm{Var}\left(\sum_i D_i\right)^{-1}\frac{1}{\sqrt{n}}\sum_i(D_i-\mathbf{E}[D_i])\right|\right|>\frac{\epsilon}{3}\right)\\
&\quad+\mathbf{P}_F^*\left(\sup_{d_F(\theta,\theta')<\delta}\left|\left|\sigma_a^n(\theta)\sigma_a^n(\theta')^{-1}-1\right|\right|\sup_{d_F(\theta,\theta')<\delta}\left|\left|\frac{1}{\sqrt{n}}\sum_i\left(U_i(\theta')-\mathbf{E}[U_i(\theta')]\right)\right|\right|>\frac{\epsilon}{3}\right).
\end{aligned}$$

Label the three terms on the right-hand side A, B, and C. Term A is bounded above by

$$\mathbf{P}_F^*\left(\sup_{d_F(\theta,\theta')<\delta}\left|\left|C\left(\frac{1}{\sqrt{n}}\sum_i(U_i(\theta)-\mathbf{E}[U_i(\theta)])-\frac{1}{\sqrt{n}}\sum_i\left(U_i(\theta')-\mathbf{E}[U_i(\theta')]\right)\right)\right|\right|>\frac{\epsilon}{3}\right)$$

for some constant C by Assumption 10. This converges to zero if we establish stochastic equicontinuity of the empirical process $\frac{1}{\sqrt{n}}\sum_i\left(U_i(\theta')-\mathbf{E}[U_i(\theta')]\right)$, which is proven below.

Term B is the product of three elements. First notice that by a uniform-in-distribution central limit theorem for random fields,[15] $n\mathrm{Var}\left(\sum_i D_i\right)^{-1}\frac{1}{\sqrt{n}}\sum_i(D_i-\mathbf{E}[D_i])=O_p(1)$ uniformly in F. Second, the moments are uniformly bounded by Assumptions 2 and 9(a), a property inherited by the element $n\mathrm{Var}\left(\sum_i D_i\right)^{-1}$. Third, Lemma B.2 implies that $\sigma_a^n(\theta)\sigma_a^n(\theta')^{-1}-1$ uniformly converges to zero as $\delta\to 0$. Then by Slutsky's theorem, $\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{F\in\mathcal{P}}\mathsf{B}=0$.

To show that C converges to zero, since $\sigma_a^n(\theta)\sigma_a^n(\theta')^{-1}-1$ uniformly converges to zero, as argued previously, by the continuous mapping theorem (van der Vaart and Wellner, 1996, Theorem 1.11.1) it is enough to show that the empirical process $\frac{1}{\sqrt{n}}\sum_i\left(U_i(\theta')-\mathbf{E}_F[U_i(\theta')]\right)$ weakly converges to a Gaussian process. Note that establishing this result also implies stochastic equicontinuity, which then implies that A converges to zero, completing the proof. Under Assumption 12, it is enough to verify the conditions of Theorem 4.9 of Arcones and Giné (1993).[16] Condition (ii) is satisfied, since the moments are uniformly bounded. To

[15] Jenish and Prucha (2009) provide a non-uniform central limit theorem for random fields, while van der Vaart and Wellner (1996) (Proposition A.5.2) provides a uniform central limit theorem for i.i.d. data.

[16] This result is not uniform in the distribution, but the extension to this case should be straight-

establish condition (iii), by Theorem 2.6.7 of van der Vaart and Wellner (1996), it suffices to show that the class of functions $\{m_a^n(\theta); \theta \in \Theta\}$ is VC-subgraph. This follows from Assumption 13 and Lemma 2.13 of Pakes and Pollard (1989),[17] which thus establishes Theorem 4.9, and hence, stochastic equicontinuity. By Theorem 2.8.2 of van der Vaart and Wellner (1996), this in turn implies pre-Gaussianity, verifying condition (i) of Theorem 4.9. ■

Proposition B.5. *Under Assumptions 2, 9, 10, 11, Assumptions A.1-A.4 of Bugni et al. (2014) hold.*

Proof. Assumption A.1 is maintained. Proposition B.4 verifies Assumption A.2. Assumption A.3 follows because the moments are uniformly bounded by Proposition B.1. Lemma B.2 implies Assumption A.4. ■

C Appendix: Proofs of Main Results

For all $n \in \mathbb{N}$, $i, j \in \mathcal{N}_n$, we abuse notation and define Γ_{ij} as the smallest set satisfying Assumption 6(a). For $A, A' \subseteq \mathcal{N}_n$, let $\Gamma(A, A') = \bigcup_{k \in A} \bigcap_{l \in A'} \Gamma_{kl}$. We will view $\Gamma(A, A')$ both as a set of positions and the set of nodes associated with these positions. Viewed as a set of nodes, let $\Gamma(A, A')^c = \mathcal{N}_n \backslash \Gamma(A, A')$.

For two sets of nodes $A, A' \subseteq \mathcal{N}_n$, let $\{A \leftrightarrow A'\} = \{\exists\, (i, j) \in A \times A' : G_{ij} = 1\}$. Let $\{A \not\leftrightarrow A'\}$ be the complement of this event. Additionally, define the event

$$\mathsf{C}^c_{A,A'} = \left\{Z_{ij} \in \mathcal{C}(||\delta_{ij}||)\ \forall (i, j) \in \Gamma(A, A') \times \Gamma(A, A')^c\right\}.^{18} \tag{22}$$

This says that realizations of Z_{ij} lies within their constraint sets $\mathcal{C}(\cdot)$, defined in Assumption

forward (cf. van der Vaart and Wellner, 1996, Theorem 2.8.3). Additionally, the theorem only applies to identically distributed random variables, whereas in our case, $\{X_i\}$ is non-identically distributed, since it depends on fixed positions ρ_i. We are not aware of weak convergence results in this case. (However, see Giné et al. (2000) and Peña and Giné (1999), Chapter 3, for some developments in this area.)

[17] In the notation of Pakes and Pollard (1989), take

$$\varphi = \sup_\theta \left| \nabla_\theta \mathbf{E}\left[\max_{g_{-A}} \max_{g_A \in \mathcal{G}_A(g_{-A}, W_A, \theta)} u' \mathbb{I}_a(g_A) f_a(Z_A^o) \,\middle|\, Z_A^o\right]\right|.$$

[18] Note that $\mathcal{C}(||\delta_{ij}||) \cap \mathcal{C}(||\delta_{ik}||) \neq \varnothing$, since $\mathcal{C}(r)$ is nondecreasing and nonempty by assumption.

4, for all indicated pairs. The next lemma states that with probability one, nodes in $\Gamma(A, A')$ are unconnected to nodes in its complement under the event $\mathsf{C}^c_{A,A'}$, provided A and A' are sufficiently distant.

Lemma C.1. *Under Assumptions 2, 4, and 6, for r sufficiently large,*

$$\sup\left\{\mathbf{P}\left(\Gamma(A,A') \leftrightarrow \Gamma(A,A')^c \,|\, \mathsf{C}^c_{A,A'}\right); A, A' \subseteq \mathcal{N}_n, d(A,A') \geqslant r, n \in \mathbb{N}\right\} = 0. \tag{23}$$

Proof. All probability statements in this proof will be conditional on the event $\mathsf{C}^c_{A,A'}$. Since $d(A,A') \geqslant r$, Assumption 6 implies that $d\big(\Gamma(A,A'), \Gamma(A,A')^c\big) \geqslant \gamma(r)$. Suppose, to obtain a contradiction, that for some $n \in \mathbb{N}$ and $A, A' \subseteq \mathcal{N}_n$, with positive probability, $\Gamma(A,A') \leftrightarrow \Gamma(A,A')^c$. Notice

$$\mathbf{P}(\Gamma(A,A') \leftrightarrow \Gamma(A,A')^c) \leqslant \max_{\substack{i \in \Gamma(A,A'),\\ j \in \Gamma(A,A')^c}} \mathbf{P}(V_{ij}(\delta_{ij}, \mathcal{E}^n_{ij}, Z_{ij}; \theta_0) \geqslant 0).$$[19]

Conditional on $\mathsf{C}^c_{A,A'}$, $Z_{ij} \in \mathcal{C}(||\delta_{ij}||)$, so it follows that the right-hand side of the above expression is bounded above by

$$\max_{\substack{i \in \Gamma(A,A'),\\ j \in \Gamma(A,A')^c}} \sup_{\mathcal{E} \in E} \sup_{z \in \mathcal{C}(||\delta_{ij}||)} \mathbf{1}\{V_{ij}(\delta_{ij}, \mathcal{E}, z; \theta_0) \geqslant 0\}.$$

But for r sufficiently large, by Assumption 6, $\gamma(r)$, and therefore $||\delta_{ij}||$, is large. Hence, by Assumption 4(b), $V_{ij}(\delta_{ij}, \mathcal{E}, z; \theta_0)$ is strictly less than zero for all $\mathcal{E} \in E$ and $z \in \mathcal{C}(||\delta_{ij}||)$, which establishes the contradiction. ∎

We next derive an asymptotic bound on the probability that the complement of $\mathsf{C}^c_{A,A'}$, denoted $\mathsf{C}_{A,A'}$, is true.

Lemma C.2. *Under Assumptions 4, 5, and 6, there exists $\kappa > 0$ such that*

$$\sup\{\mathbf{P}(\mathsf{C}_{A,A'}); A, A' \subseteq \mathcal{N}_n, d(A,A') \geqslant r, n \in \mathbb{N}\} = O\left(\min\{|A|, |A'|\} r^{-3d_\rho - \kappa}\right).$$

Proof. Let $A, A' \subseteq \mathcal{N}_n$ and $d(A,A') \geqslant r$. By the union bound,

$$\mathbf{P}(\mathsf{C}_{A,A'}) \leqslant |A| \max_{k \in A} \mathbf{P}\left(\exists\, i \in \cap_{l \in A'} \Gamma_{kl}, j \in \Gamma(A,A')^c : Z_{ij} \notin \mathcal{C}(||\delta_{ij}||)\right). \tag{24}$$

Fix $k \in A$, and let $\min_{l \in A'} ||\delta_{kl}|| = \tilde{r}$. Then Assumption 6 implies that, for any $i \in \cap_{l \in A'} \Gamma_{kl}$,

[19] For simplicity, we are assuming that the network is undirected.

$\min\{||\delta_{ij}||; j \in \Gamma(A, A')^c\} \geqslant \gamma(\tilde{r})$. This motivates us to define for each $i \in \cap_{l\in A'}\Gamma_{kl}$ a partition of $\Gamma(A, A')^c$ consisting of sets $\{S_{i,\tilde{r},t}\}_{t=0}^{\infty}$ such that for each t, $S_{i,\tilde{r},t} = \{j \in \mathcal{N}_n : ||\delta_{ij}|| \in (\gamma(\tilde{r}) + t, \gamma(\tilde{r}) + t + 1]\} \cap \Gamma(A, A')^c$. Since $\mathcal{C}$ is nondecreasing (Assumption 4),

$$\begin{aligned}\mathbf{P}\Big(\exists\, i \in \cap_{l\in A'}\Gamma_{kl}, j \in&\Gamma(A, A')^c : Z_{ij} \notin \mathcal{C}(||\delta_{ij}||)\Big) \\ &\leqslant \mathbf{P}\left(\bigcup_{t=0}^{\infty}\{\exists\, i \in \cap_{l\in A'}\Gamma_{kl}, j \in S_{i,\tilde{r},t} : Z_{ij} \notin \mathcal{C}(\gamma(\tilde{r}) + t)\}\right) \\ &\leqslant \sum_{t=0}^{\infty} |\cap_{l\in A'}\Gamma_{kl}| \max_{\substack{i\in\cap_{l\in A'}\Gamma_{kl} \\ j\in S_{i,\tilde{r},t}}} |S_{i,\tilde{r},t}|\, \mathbf{P}\left(Z_{ij} \notin \mathcal{C}(\gamma(\tilde{r}) + t)\right). \qquad (25)\end{aligned}$$

By Assumption 6(a), $\cap_{l\in A'}\Gamma_{kl} \subseteq \cap_{l\in A'}\mathcal{B}_k^n(||\delta_{kl}||)$, which has cardinality $O(\tilde{r}^{d_\rho})$ by the construction of k, Assumption 5, and Lemma A.1 of Jenish and Prucha (2009). This lemma also implies that $\sup\{|S_{i,\tilde{r},t}|; i \in \cap_{l\in A'}\Gamma_{kl}\} = O((\gamma(\tilde{r})+t)^{d_\rho-1})$. Furthermore, by Assumption 4(a), $\mathbf{P}\left(Z_{ij} \notin \mathcal{C}(\gamma(\tilde{r}) + t)\right)$ is uniformly $O(\tau(\gamma(\tilde{r})+t))$. Combining these facts, the summands of (25) are of asymptotic order

$$\tilde{r}^{d_\rho}(\gamma(\tilde{r}) + t)^{d_\rho-1}\tau(\gamma(\tilde{r}) + t) \leqslant \tilde{r}^{d_\rho}(\gamma(\tilde{r}) + t)^{d_\rho+\beta}\tau(\gamma(\tilde{r}) + t)t^{-1-\beta}, \qquad (26)$$

for arbitrary $\beta \in (0, \varphi)$ (φ defined in Assumption 4(a)) and r sufficiently large, using the fact that $\gamma(\tilde{r})$ is diverging with $\tilde{r}$. Notice that for t sufficiently large, by Assumption 4(a), there exists $K > 0$ such that

$$\begin{aligned}(\gamma(\tilde{r}) + t)^{d_\rho+\beta}\tau(\gamma(\tilde{r}) + t) &= (\gamma(\tilde{r}) + t)^{-4d_\rho-\varphi+\beta}\underbrace{\tau(\gamma(\tilde{r}) + t)(\gamma(\tilde{r}) + t)^{5d_\rho+\varphi}}_{\to c>0} \\ &\leqslant K\gamma(\tilde{r})^{-4d_\rho-\varphi+\beta}\tau(\gamma(\tilde{r}))\gamma(\tilde{r})^{5d_\rho+\varphi} \\ &= K\gamma(\tilde{r})^{d_\rho+\beta}\tau(\gamma(\tilde{r})) \\ &\leqslant K\mu\tilde{r}^{-4d_\rho-\varphi}\gamma(\tilde{r})^{\beta} \qquad \text{(Assumption 6(b))} \\ &< K\mu\tilde{r}^{-4d_\rho-(\varphi-\beta)}. \qquad (\gamma(\tilde{r}) < \tilde{r})\end{aligned}$$

Since $\sum_{t=0}^{\infty} t^{-1-\beta}$ converges, (26) and the above argument imply that (25) $= O(\tilde{r}^{-3d_\rho-\kappa})$ for $\kappa = \varphi - \beta$. Since $d(A, A') \geqslant r$ implies $\tilde{r} \geqslant r$, and $\varphi > \beta$, the claim follows from (24). ∎

The next lemma shows that, for any n, $(\psi_i^n; i \in A) \perp\!\!\!\perp (\psi_j^n; j \in A')$ conditional on $\mathsf{C}_{A,A'}^c$. For $A \subseteq \mathcal{N}_n$, define $\psi_A^n = (\psi_i^n; i \in A)$.

Lemma C.3. *Under Assumptions 1, 3, and 7, for any $H \in \sigma(\psi_A^n)$ and $H' \in \sigma(\psi_{A'}^n)$ and r sufficiently large,*

$$\sup\big\{\big|\mathbf{P}(\psi_A^n \in H \cap \psi_{A'}^n \in H' \mid \mathsf{C}_{A,A'}^c) - \mathbf{P}(\psi_A^n \in H \mid \mathsf{C}_{A,A'}^c)\mathbf{P}(\psi_{A'}^n \in H' \mid \mathsf{C}_{A,A'}^c)\big|; \\ A, A' \subseteq \mathcal{N}_n, d(A,A') \geqslant r, n \in \mathbb{N}\big\} = 0. \quad (27)$$

Proof. Fix $A, A' \subseteq \mathcal{N}_n$, $d(A,A') \geqslant r$, and $n \in \mathbb{N}$. Define

$$\lambda_{\Gamma(A,A')} = \big(\lambda_S(W_S, \nu_S; \theta_0); S \in \mathcal{S}(W,\theta_0), S \subseteq \Gamma(A,A')\big),$$
$$\nu_{\Gamma(A,A')} = \big(\nu_S; S \in \mathcal{S}(W,\theta_0), S \subseteq \Gamma(A,A')\big).$$

In other words, $\nu_{\Gamma(A,A')}$ is the vector of public signals ν_S associated with each isolated society $S \subseteq \Gamma(A,A')$, while $\lambda_{\Gamma(A,A')}$ is the vector of subnetwork selection mechanisms for each of these isolated societies. Similarly define $\lambda_{\Gamma(A,A')^c}$ and $\nu_{\Gamma(A,A')^c}$.

Since $A \subseteq \Gamma(A,A')$, $A' \subseteq \Gamma(A,A')^c$, and (23) holds by Lemma C.1, $C_i(G) \subseteq \Gamma(A,A')$ and $C_j(G) \subseteq \Gamma(A,A')^c$ for all $i \in A$ and $j \in A'$ with probability one conditional on $\mathsf{C}_{A,A'}^c$. Then conditional on $\mathsf{C}_{A,A'}^c$, by the definition of node statistics, the event $\{\psi_A^n \in H \cap \psi_{A'}^n \in H'\}$ is equivalent to the event

$$\big\{G_{\Gamma(A,A')} \in \mathsf{G} \cap G_{-\Gamma(A,A')} \in \mathsf{G}' \cap W_{\Gamma(A,A')} \in \mathsf{W} \cap W_{\Gamma(A,A')^c} \in \mathsf{W}'\big\} \quad (28)$$

for some $\mathsf{G} \in \sigma(G_{\Gamma(A,A')})$, $\mathsf{G}' \in \sigma(G_{-\Gamma(A,A')})$, $\mathsf{W} \in \sigma(W_{\Gamma(A,A')})$, and $\mathsf{W}' \in \sigma(W_{\Gamma(A,A')^c})$. Thus,

$$\mathbf{P}(\psi_A^n \in H \cap \psi_{A'}^n \in H' \mid \mathsf{C}_{A,A'}^c) = \mathbf{E}\big[\mathbf{P}\big(G_{\Gamma(A,A')} \in \mathsf{G} \cap G_{-\Gamma(A,A')} \in \mathsf{G}' \\ \cap W_{\Gamma(A,A')} \in \mathsf{W} \cap W_{\Gamma(A,A')^c} \in \mathsf{W}' \mid W \in \mathcal{W}, \mathsf{C}_{A,A'}^c\big) \;\big|\; \mathsf{C}_{A,A'}^c\big]. \quad (29)$$

Henceforth, purely for ease of notation, we will only consider the case in which node statistics ψ_i^n do not depend on W, so that we can drop the event $\{W_{\Gamma(A,A')} \in \mathsf{W} \cap W_{\Gamma(A,A')^c} \in \mathsf{W}'\}$ from (29). With these changes, by Assumption 7(b)

$$(29) = \mathbf{E}\left[\sum_{g \in \mathsf{G}\times\mathsf{G}'} \mathbf{P}\left(g \in \mathcal{G}_{\mathcal{N}_n}(W,\theta_0) \cap \lambda_{\mathcal{N}_n}(W,\nu,\theta_0) = g \mid W, \mathsf{C}_{A,A'}^c\right) \;\middle|\; \mathsf{C}_{A,A'}^c\right], \quad (30)$$

where by "$g \in \mathsf{G}\times\mathsf{G}'$" we mean "$g : g_{\Gamma(A,A')} \in \mathsf{G}, g_{-\Gamma(A,A')} \in \mathsf{G}'$." By (23) and Assumption 3, conditional on $\{W, \mathsf{C}^c_{A,A'}\}$, $g \in \mathcal{G}_{\mathcal{N}_n}(W, \theta_0)$ if and only if

$$g_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')}, \theta_0),$$
$$g_{\Gamma(A,A')^c} \in \mathcal{G}_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c}, \theta_0),$$

and $V_{ij}(g, W; \theta_0) < 0$ for all $i \in \Gamma(A, A')$ and $j \in \Gamma(A, A')^c$ (provided r is sufficiently large). Then by Assumption 7(c), $\mathbf{P}\big(g \in \mathcal{G}_{\mathcal{N}_n}(W, \theta_0) \cap \lambda_{\mathcal{N}_n}(W, \nu, \theta_0) = g \,|\, W, \mathsf{C}^c_{A,A'}\big)$ equals

$$\begin{aligned}\mathbf{P}\big(g_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')}, \theta_0) \cap g_{\Gamma(A,A')^c} \in \mathcal{G}_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c}, \theta_0)&\\ \cap\, \lambda_{\Gamma(A,A')}(W_{\Gamma(A,A')}, \nu_{\Gamma(A,A')}, \theta_0) = g_{\Gamma(A,A')}&\\ \cap\, \lambda_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c}, \nu_{\Gamma(A,A')^c}, \theta_0) = g_{\Gamma(A,A')^c}&\\ \cap\, V_{ij} < 0 \text{ for all } i \in \Gamma(A, A') \text{ and } j \in \Gamma(A, A')^c \,|\, W, \mathsf{C}^c_{A,A'}\big).&\end{aligned} \tag{31}$$

We can omit the last event in the following expressions, since it holds with probability one under $\mathsf{C}^c_{A,A'}$ by (23). By construction, $\nu_S \perp\!\!\!\perp W$, and $\{\nu_S\}_S$ is independently distributed by Assumption 1, so (31) equals

$$\begin{aligned}&\mathbf{P}\big(g_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')}, \theta_0)\\ &\qquad \cap\, \lambda_{\Gamma(A,A')}(W_{\Gamma(A,A')}, \nu_{\Gamma(A,A')}, \theta_0) = g_{\Gamma(A,A')} \,|\, W, \mathsf{C}^c_{A,A'}\big)\\ &\times \mathbf{P}\big(g_{\Gamma(A,A')^c} \in \mathcal{G}_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c}, \theta_0)\\ &\qquad \cap\, \lambda_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c}, \nu_{\Gamma(A,A')^c}, \theta_0) = g_{\Gamma(A,A')^c} \,|\, W, \mathsf{C}^c_{A,A'}\big).\end{aligned} \tag{32}$$

We next argue that

$$\{S \in \mathcal{S}(W, \theta_0) : S \subseteq \Gamma(A, A')\} \perp\!\!\!\perp W_{\Gamma(A,A')^c} \,|\, \mathsf{C}^c_{A,A'}. \tag{33}$$

Fix any S in this set. As argued in (31), (23) implies that $G_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')}, \theta_0)$, which is only random via $W_{\Gamma(A,A')}$. Since $S \subseteq \Gamma(A, A')$, by the definition of isolated societies, its elements are fully determined by $\mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')}, \theta_0)$. Furthermore, $\mathsf{C}^c_{A,A'}$ only restricts the marginal distributions of $\{X_i\}_{i=1}^n$ and $\{\zeta_{ij}\}_{i,j}$, so Assumption 1 implies $W_{\Gamma(A,A')} \perp\!\!\!\perp W_{\Gamma(A,A')^c} \,|\, \mathsf{C}^c_{A,A'}$.[20] The claim then follows.

[20]The claim that $\mathsf{C}^c_{A,A'}$ only restricts the marginal distributions of attributes follows from the construction of $\mathcal{C}$ as the Cartesian product of set-valued functions, one for each attribute. See Assumption 4.

Equation (33) implies that

$$\begin{aligned}&\mathbf{P}\big(g_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')},\theta_0)\\&\qquad\qquad \cap\, \lambda_{\Gamma(A,A')}(W_{\Gamma(A,A')},\nu_{\Gamma(A,A')},\theta_0) = g_{\Gamma(A,A')} \,|\, W, \mathsf{C}^c_{A,A'}\big)\\&= \mathbf{P}\big(g_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')},\theta_0)\\&\qquad\qquad \cap\, \lambda_{\Gamma(A,A')}(W_{\Gamma(A,A')},\nu_{\Gamma(A,A')},\theta_0) = g_{\Gamma(A,A')} \,|\, W_{\Gamma(A,A')}, \mathsf{C}^c_{A,A'}\big)\end{aligned}$$

and similarly for the second conditional probability in (32) Combining these facts, (30) equals

$$\begin{aligned}\mathbf{E}\Bigg[&\sum_{g_{\Gamma(A,A')}\in \mathsf{G}} \mathbf{P}\big(g_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')},\theta_0)\\&\qquad\qquad \cap\, \lambda_{\Gamma(A,A')}(W_{\Gamma(A,A')},\nu_{\Gamma(A,A')},\theta_0) = g_{\Gamma(A,A')} \,|\, W_{\Gamma(A,A')}, \mathsf{C}^c_{A,A'}\big)\\\times &\sum_{g_{\Gamma(A,A')^c}\in \mathsf{G}'} \mathbf{P}\big(g_{\Gamma(A,A')^c} \in \mathcal{G}_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c},\theta_0)\\&\qquad \cap\, \lambda_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c},\nu_{\Gamma(A,A')^c},\theta_0) = g_{\Gamma(A,A')^c} \,|\, W_{\Gamma(A,A')^c}, \mathsf{C}^c_{A,A'}\big) \,\Big|\, \mathsf{C}^c_{A,A'}\Bigg].\end{aligned} \tag{34}$$

Since, as argued previously, $W_{\Gamma(A,A')} \perp\!\!\!\perp W_{\Gamma(A,A')^c} \,|\, \mathsf{C}^c_{A,A'}$, (34) equals

$$\begin{aligned}&\sum_{g_{\Gamma(A,A')}\in \mathsf{G}} \mathbf{P}\big(g_{\Gamma(A,A')} \in \mathcal{G}_{\Gamma(A,A')}(W_{\Gamma(A,A')},\theta_0)\\&\qquad\qquad \cap\, \lambda_{\Gamma(A,A')}(W_{\Gamma(A,A')},\nu_{\Gamma(A,A')},\theta_0) = g_{\Gamma(A,A')} \,|\, \mathsf{C}^c_{A,A'}\big)\\\times &\sum_{g_{\Gamma(A,A')^c}\in \mathsf{G}'} \mathbf{P}\big(g_{\Gamma(A,A')^c} \in \mathcal{G}_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c},\theta_0)\\&\qquad\qquad \cap\, \lambda_{\Gamma(A,A')^c}(W_{\Gamma(A,A')^c},\nu_{\Gamma(A,A')^c},\theta_0) = g_{\Gamma(A,A')^c} \,|\, \mathsf{C}^c_{A,A'}\big).\end{aligned}$$

By (23) and Assumption 7, this equals the right-hand side of (27) using the arguments for (29) and (30), as desired. ■

Proof of Theorem 1. Let $A, A' \subseteq \mathcal{N}_n$, $H \in \sigma(\psi_i^n; i \in A)$, and $H' \in \sigma(\psi_i^n; i \in A')$. Define $\mathsf{C}^c_{A,A'}$ and its complement $\mathsf{C}_{A,A'}$ as in (22). Then

$$\begin{aligned}|\mathbf{P}(H\cap H') - \mathbf{P}(H)\mathbf{P}(H')| &\leqslant \big|\mathbf{P}(H\cap H' \,|\, \mathsf{C}_{A,A'}) - \mathbf{P}(H \,|\, \mathsf{C}_{A,A'})\mathbf{P}(H' \,|\, \mathsf{C}_{A,A'})\big|\, \mathbf{P}(\mathsf{C}_{A,A'})\\&+ \big|\mathbf{P}(H\cap H' \,|\, \mathsf{C}^c_{A,A'}) - \mathbf{P}(H \,|\, \mathsf{C}^c_{A,A'})\mathbf{P}(H' \,|\, \mathsf{C}^c_{A,A'})\big|\, \mathbf{P}(\mathsf{C}^c_{A,A'}) + 2\mathbf{P}(\mathsf{C}_{A,A'}).\end{aligned} \tag{35}$$

Let $\alpha_n(A, A' \mid \mathsf{C}^c_{A,A'})$ be the conditional mixing coefficient, which is defined analogously to $\alpha_n(A, A')$, replacing the probabilities in the definition with conditional probabilities, the conditioning event being $\{\mathsf{C}^c_{A,A'}\}$. Then by (35),

$$\alpha_n(A, A') \leqslant 3\mathbf{P}(\mathsf{C}_{A,A'}) + \alpha_n(A, A' \mid \mathsf{C}^c_{A,A'}). \tag{36}$$

Next define

$$\bar{\alpha}_{a,a'}(r \mid \mathsf{C}^c_r) = \sup\{\alpha_n(A, A' \mid \mathsf{C}^c_{A,A'}); |A| \leqslant a, |A'| \leqslant a', d(A, A') \geqslant r, n \in \mathbb{N}\}, \quad \text{and}$$
$$\mathsf{c}_{a,a'}(r) = \sup\{\mathbf{P}(\mathsf{C}_{A,A'}); |A| \leqslant a, |A'| \leqslant a', d(A, A') \geqslant r, n \in \mathbb{N}\}.$$

Then by (36),

$$\bar{\alpha}_{a,a'}(r) \leqslant 3\mathsf{c}_{a,a'}(r) + \bar{\alpha}_{a,a'}(r \mid \mathsf{C}^c_r).$$

The coefficient $\bar{\alpha}_{a,a'}(r \mid \mathsf{C}^c_r)$, is zero with probability one for sufficiently large r by Lemma C.3. By Lemma C.2, $\mathsf{c}_{a,a'}(r) \leqslant \min\{a, a'\}\hat{\alpha}(r)$ with $\hat{\alpha}(r) = O(r^{-3d_\rho-\kappa})$ for some $\kappa > 0$. Hence, Condition 1 holds with $\eta = 1$. ■

PROOF OF THEOREM 2. L_1 convergence follows from Theorem 3 of Jenish and Prucha (2009), since Theorem 1 implies $\sum_{r=1}^{\infty} r^{d_\rho-1}\bar{\alpha}_{1,1}(r)^{1/2} \to 0$. Asymptotic normality follows from Theorem A.1 of Jenish and Prucha (2012). Assumption 5 is a weaker version of their increasing domain assumption that allows for finitely many nodes to have the same position. This does not alter the conclusion of their Lemma A.1, nor, therefore, the aforementioned theorems. ■

PROOF OF PROPOSITION 3.1. For a given a, let $A \subseteq \mathcal{N}_n$ with $|A| = a$ and $f_a \in \mathcal{F}_A$. Define

$$Q^n_A(W, f_a, \theta) = \{\mathbb{I}_a(g_A) f_A(Z^o_A) : \exists\, g_{-A} \text{ such that } (g_A, g_{-A}) \in \mathcal{G}_{\mathcal{N}_n}(W, \theta)\},$$
$$\mathbb{E}_\theta\left[Q^n_A(W, f_a, \theta)\right] = \{\mathbf{E}_\theta[q] : q \in Q^n_A(W, f_a, \theta)\}.$$

By definition, $\theta \in \Theta_I$ if and only if for all n, $a > 1$, and $f_a \in \mathcal{F}_a$, there exists some $E(q) \in \mathbb{E}_\theta\left[Q^n_A(W, f_a, \theta)\right]$ such that

$$\lim_{n\to\infty} \frac{1}{nM_n} \sum_{|A|=a} \big(\mathbf{E}[\mathbb{I}_a(G_A) f_a(Z^o_A)] - E(q)\big) = 0.$$

This implies that

$$\lim_{n\to\infty} \frac{1}{nM_n} \sum_{|A|=a} \left[u'\mathbf{E}[\mathbb{I}_a(g_A) f_A(Z_A^o)] - u'E(q)\right] \leqslant 0 \quad \forall u \in \mathbb{R}^{d_a}$$

$$\Rightarrow \limsup_{n\to\infty} \frac{1}{nM_n} \sum_{|A|=a} \left[u'\mathbf{E}[\mathbb{I}_a(g_A) f_A(Z_A^o)] - \sup_{E(q)\in\mathbb{E}_\theta[Q_A^n(W,f_a,\theta)]} u'E(q) \right] \leqslant 0 \quad \forall u \in \mathbb{R}^{d_a}. \tag{37}$$

Since for any n, A, and X, the set $Q_A^n(W, f_a, \theta)$ contains at most $2^{|\mathbf{G}^{\leftrightarrow}_{u,a}|} \times |\text{Range}(\mathcal{F}_a)|$ values, and $\text{Range}(\mathcal{F}_a) = \{0, 1\}$ by assumption, it is a random closed set (Molchanov, 2005). Thus, by Theorem 2.1.47(iv) of Molchanov (2005),

$$\sup_{E(q)\in\mathbb{E}_\theta[Q_A^n(W,f_a,\theta)]} u'E(q) = \mathbf{E}_\theta \left[\sup_{q\in Q_A^n(W,f_a,\theta)} u'q \right].$$

Hence, (37) holds if and only if

$$\limsup_{n\to\infty} \frac{1}{nM_n} \sum_{|A|=a} \left(u'\mathbf{E}[\mathbb{I}_a(g_A) f_A(Z_A^o)] - \mathbf{E}_\theta \left[\sup_{q\in Q_A^n(W,f_a,\theta)} u'q \right] \right) \leqslant 0 \quad \forall u \in \mathbb{R}^{d_a},$$

$$\Leftrightarrow \limsup_{n\to\infty} \frac{1}{nM_n} \sum_{|A|=a} \left(u'\mathbf{E}[\mathbb{I}_a(g_A) f_A(Z_A^o)] - \mathbf{E}_\theta \left[\max_{g\in\mathcal{G}_{\mathcal{N}_n}(W,\theta)} u'\mathbb{I}_a(g_A) f_A(Z_A^o) \right] \right) \leqslant 0,$$

which implies

$$\limsup_{n\to\infty} \frac{1}{nM_n} \sum_{|A|=a} \left(u'\mathbf{E}[\mathbb{I}_a(g_A) f_A(Z_A^o)] - \mathbf{E}_\theta \left[\max_{g_{-A}} \max_{g_A\in\mathcal{G}_A(g_{-A},W,\theta)} u'\mathbb{I}_a(g_A) f_A(Z_A^o) \right] \right) \leqslant 0,$$

for all $u \in \mathbb{R}^{d_a}$. ∎

References

Agarwal, N. and W. Diamond, "Identification and Estimation in Two-Sided Matching Markets," *working paper*, 2014. 7

Andrews, D. and X. Shi, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666. 28, 34

Arcones, M. and E. Giné, "Limit Theorems for U-Processes," *The Annals of Probability*, 1993, *21* (3), 1494–1542. 49

Badev, A., "Discrete Games in Endogenous Networks: Theory and Policy," *working paper*, 2013. 2

Bajari, P., H. Hong, and S. Ryan, "Identification and Estimation of Discrete Games of Complete Information," *Econometrica*, 2010, *78* (5), 1529–1568. 7

Beresteanu, A., I. Molchanov, and F. Molinari, "Sharp Identification Regions in Models with Convex Moment Inequalities," *Econometrica*, 2011, *79* (6), 1785–1821. 7, 29, 32, 33, 34

Bhamidi, S., G. Bresler, and A. Sly, "Mixing Time of Exponential Random Graphs," *The Annals of Applied Probability*, 2011, *21* (6), 2146–2170. 6

Bisin, A., A. Moro, and G. Topa, "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions," *working paper*, 2011. 7

Boucher, V. and I. Mourifié, "My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks," *working paper*, 2013. 5, 12, 26

Bramoullé, Y. and B. Fortin, "Social Networks: Econometrics," in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics.*, Basingstoke: Palgrave Macmillan, 2010. 3

Bresnahan, T. and P. Reiss, "Entry in Monopoly Markets," *Review of Economic Studies*, 1990, *57* (4), 387–424. 7

Bugni, F., I. Canay, and X. Shi, "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," *working paper*, 2014. 34, 45, 50

Chandrasekhar, A., "Econometrics of Network Formation," *working paper*, 2014. 15

— **and M. Jackson**, "Tractable and Consistent Random Graph Models," *working paper*, 2014. 6

Chernozhukov, V., H. Hong, and E. Tamer, "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 2007, *75* (5), 1243–1284. 34

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman, "An Empirical Model for Strategic Network Formation," *working paper*, 2010. 6, 12

Ciliberto, F. and E. Tamer, "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 2009, *77* (6), 1791–1828. 7, 34

Comola, M., "Estimating Local Network Externalities," *working paper*, 2012. 3

DeWeert, J., "Risk-Sharing and Endogenous Network Formation," in S. Dercon, ed., *Insurance against Poverty*, Oxford University Press,, 2004, chapter 10, pp. 297–216. 3

Dzemski, A., "An Empirical Model fo Dyadic Link Formation in a Network with Unobserved Heterogeneity," *working paper*, 2014. 5

Fafchamps, M. and F. Gubert, "The Formation of Risk-Sharing Networks," *Journal of Economic Development*, 2007, *83*, 326–350. 3

Fox, J., "Estimating Matching Games with Transfers," *working paper*, 2010. 7

Frees, E., "Infinite Order U-Statistics," *Scandinavian Journal of Statistics*, 1989, *16* (1), 29–45. 27

Galichon, A. and M. Henry, "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 2011, *78* (4), 1264–1298. 7, 29

Giné, E., R. Latala, and J. Zinn, "Exponential and Moment Inequalities for *U*-Statistics," in E. Giné, D. Mason, and J. Wellner, eds., *High Dimensional Probability II*, Vol. 47, New York: Springer, 2000, chapter 2, pp. 13–38. 50

Goldsmith-Pinkham, P. and G. Imbens, "Social Networks and the Identification of Peer Effects," *Journal of Business and Economic Statistics*, 2013, *31* (3), 253–264. 2, 12

Graham, B., "Comment: 'Social Networks and the Identification of Peer Effects'," *Journal of Business and Economic Statistics*, 2013, *31* (3), 266–270. 12

_ , "An Empirical Model of Network Formation: Detecting Homophily when Agents Are Heterogeneous," *working paper*, 2014. 5, 15

Hájek, J., "Asymptotic Normality of Simple Linear Rank Statistics Under Alternatives," *Annals of Mathematical Statistics*, 1968, *39* (2), 325–346. 40

Henry, M., R. M/'eango, and M. Queyranne, "Combinatorial Bootstrap Inference in Partially Identified Incomplete Structural Models," *Quantitative Economics*, forthcoming. 7

Hsieh, C. and L. Lee, "A Structural Modeling Approach for Network Formation and Social Interactions—with Applications to Students' Friendship Choices and Selectivity on Activities," *working paper*, 2012. 6

Jackson, M., *Social and Economic Networks*, Princeton University Press, 2008. 12

Jenish, N., "Spatial Semiparametric Model with Endogenous Regressors," *working paper*, 2013. 34, 43

_ **and I. Prucha**, "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields," *Journal of Econometrics*, 2009, pp. 86–98. 14, 18, 38, 44, 49, 52, 56

_ **and** _ , "On spatial processes and asymptotic inference under near-epoch dependence," *Journal of Econometrics*, 2012, *170*, 178–190. 56

Jong, P. De, "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 1987, *75*, 261–277. 46

Lee, A., U-*Statistics*, Marcel Dekker, Inc., 1990. 40

Leung, M., "Two-Step Estimation of Network-Formation Models with Incomplete Information," *working paper*, 2014. 5

McPherson, M., L. Smith-Lovin, and J. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 2001, *27*, 415–444. 13

Mele, A., "A Structural Model of Segregation in Social Networks," *working paper*, 2013. 6, 12, 26

Menzel, K., "Inference for Games with Many Players," *working paper*, 2014. 7

_ , "Large Matching Markets as Two-Sided Demand Systems," *working paper*, 2014. 7

Miyauchi, Y., "Structural Estimation of a Pairwise Stable Network Formation with Non-negative Externality," *working paper*, 2013. 5, 12, 32

Molchanov, I., *Theory of Random Sets*, London: Springer-Verlag, 2005. 57

Myerson, R., "Graphs and Cooperation in Games," *Mathematics of Operations Research*, 1977, *2*, 225–229. 8

na, V. Pe and E. Giné, *Decoupling: From Dependence to Independence*, Spring, 1999. 50

Pakes, A. and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 1989, *57* (5), 1027–1057. 50

— **, J. Porter, K. Ho, and J. Ishii**, "Moment Inequalities and Their Application," *working paper*, 2011. 34

Powell, W., D. White, K. Koput, and J. Owen-Smith, "Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences," *American Journal of Sociology*, 2005, *110* (4), 1132–1205. 3

Romano, J. and A. Shaikh, "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 2008, *138*, 2786–2807. 34

Shalizi, C. and A. Thomas, "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods and Research*, 2011, *40* (2), 211–239. 2

Sheng, S., "Identification and Estimation of Network Formation Games," *working paper*, 2014. 5, 8, 12, 21, 26, 27, 32

Song, K., "Econometric Inference on a Large Bayesian Game," *working paper*, 2012. 7

Storn, R. and K. Price, "Differential Evolution: A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," *Journal of Global Optimization*, 1997, *11* (4), 341–359. 34

Tamer, E., "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 2003, *70* (1), 147–165. 7

— , "Partial Identification in Econometrics," *Annual Reviews of Economics*, 2010, *2*, 167–195. 29

Uetake, K., "Strategic Network Formation and Performance in the Venture Capital Industry," *working paper*, 2012. 2

van der Vaart, A. and J. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics* Springer Series in Statistics, Springer, 1996. 49, 50

Wan, Y., "An Integration-Based Approach to Moment Inequality Models," *working paper*, 2013. 34