

# A WEAK LAW FOR MOMENTS OF PAIRWISE-STABLE NETWORKS

Michael P. Leung\*

January 2, 2016

ABSTRACT. We develop asymptotic theory for strategic network-formation models under the assumption that the econometrician observes a single large pairwise-stable network. Drawing on new techniques in the random-graphs literature, we derive sufficient conditions for an unconditional weak law of large numbers for a useful class of network moments. Under these conditions, the model generates realistic networks that are sparse and may contain “giant” connected subnetworks, two well-known properties of real-world social networks. The conditions also conveniently suggest a new method to simulate counterfactual networks that avoids a well-known curse of dimensionality. Lastly, we characterize the identified set of structural parameters based on a tractable class of pair-level network moments and construct consistent estimators.

---

\*M.I.T., Institute for Data, Systems, and Society. E-mail: mpleung@mit.edu. I thank seminar audiences at Duke, Penn State, and UNC, as well as 2015 participants of the California Econometrics Conference at USC, MIT Econometrics Lunch, MIT Social Analytics Lab, and WIN Workshop at NYU.

# 1 Introduction

There has been a surge of recent interest in statistical methods for network data. A growing econometric literature studies “strategic” models of network formation, which are extensions of dyadic link-formation models (Bramoullé and Fortin, 2010; Fafchamps and Gubert, 2007) that allow for *network externalities*, or endogenous network-dependent regressors. Strategic models easily deliver “network-effects” parameters, for example *preferential attachment*, the partial-equilibrium increase in the likelihood of link formation due to an exogenous increase in the alter’s degree.<sup>1</sup> These models are also useful for simulating counterfactual networks to study the effect of policy interventions on network structure. For instance, policies that reallocate peers across classrooms may have little effect if high-ability peers self-segregate within classrooms (Carrell et al., 2013). This motivates the use of structural models of peer-group formation to simulate the effect of reallocation policies on peer diversity within groups.

It is challenging to estimate strategic models of network formation for several reasons. First, inference requires a large sample of sufficiently uncorrelated observations, but network externalities generate statistical dependence between network links. Moreover, in practice, the econometrician typically observes only a small number of plausibly independent networks. Second, as is well-known in the empirical games literature, strategic models are typically incompletely specified due to the existence of multiple equilibria (pairwise-stable networks), which creates new challenges for inference (de Paula, 2013). Third, the number of possible networks on a set of  $n$  nodes is enormous, on the order of  $e^{n^2}$ , which often leads to a computational curse of dimensionality when it comes simulating counterfactual networks or even estimating models.

A fourth challenge is that little is known about the ability of strategic models to generate networks with structural properties that match those of real-world social networks. Two predominant features of such networks are *sparsity* and *percolation* (Barabási, 2015, Chapter 3). Sparsity is the requirement that the expected degree is of much smaller order than  $n - 1$ , the number of potential links in a network of  $n$  nodes. For instance, the scientific collaboration network discussed in Barabási (2015) has over 20,000 nodes, but the average degree is eight. Percolation is the existence of a *giant component*, which is formally defined as a component whose size is of asymptotic order  $n$ .<sup>2</sup> This corresponds to the observation that social networks tend to feature a “large” connected subnetwork. While several papers in the econometric literature provide conditions for the emergence of sparse networks, conditions for percolation are only known for random-graph models that lack network externalities.<sup>3</sup>

---

<sup>1</sup>The *degree* of a node is the number of links she forms.

<sup>2</sup>The *components* of a network are maximal subnetworks such that every pair in the subnetwork is *path-connected*. Two nodes  $i$  and  $j$  are path-connected in a network if there exists a path from  $i$  to  $j$ . A *path* in a network from node  $i$  to  $j$  is a distinct sequence of nodes starting with  $i$  and ending with  $j$  such that for each  $k, k'$  in this sequence,  $k$  and  $k'$  are directly linked in the network.

<sup>3</sup>Note that that it is the dual requirement of sparsity and percolation that makes the problem

## A WEAK LAW FOR NETWORK MOMENTS

At first glance, the possibility of statistical inference appears incompatible with percolation. The challenge is that even local network externalities can generate global dependence among network links. That is, even if link formation between  $i$  and  $j$  only depends on their direct links, because link formation between  $j$  and  $k$  depends on the direct links formed by  $k$ , it follows that node  $i$ 's links depends indirectly on node  $k$ 's links, and by induction, on links farther in the network, as well. A consequence of these indirect dependencies is that the perturbation of a single link may trigger a global cascade of link alterations. Hence, if a network percolates, so that most nodes lie within a single giant component, it may appear that links within this component should be highly statistically dependent.

A key insight of this paper is that node pairs that draw large random-utility shocks form links regardless of the state of the network and therefore act as barriers to link-alteration cascades if externalities are local. We say that such links are *exogenously realized*. For example, if two subnetworks can only be connected by a single path of links, and some pair of nodes on this path forms their link exogenously, perturbations of links in one subnetwork do not spill over into the second subnetwork. Thus, we argue that conditional on a link forming exogenously on this path, the two subnetworks are independent. In general, we show that if a large enough share of links forms exogenously, a network can be partitioned into conditionally independent subunits even if a giant component exists. This result forms the basis of our proof of a law of large numbers for network moments.

Our specific contributions are as follows.

1. (LLN) We develop conditions under which network dependence is limited and a weak law of large numbers holds for a useful class of network moments. This result is derived as a corollary of a general weak law for functionals of a large class of random graphs, which may be of independent interest. We can characterize the unconditional limit of network moments as expectations of their analogs applied to an appropriate Poisson limit model. The derivation of explicit limiting constants is new to the literature and obtained through the construction of a partial coupling between the finite model, which generates the observed network, and the Poisson limit. The general weak law extends a result due to [Penrose and Yukich \(2003\)](#), and the coupling is based on their construction.
2. (Realistic Networks) A key idea of the paper is that strategic models can often be viewed as link-formation processes on a random graph, models whose percolative properties are well known. We can then draw on results in random-graph theory to show that our model rationalizes sparsity and to derive conditions for percolation.
3. (Counterfactuals) A naive method of simulating pairwise-stable networks is to

---

challenging, as dense graphs have expected degree of order  $n$ , which trivially leads to fully connected networks.

iterate through all possible networks and check stability. This method is infeasible even for moderately large networks since the number of networks is exponential in the number of nodes.<sup>4</sup> Our assumptions conveniently suggest a feasible procedure for simulating pairwise-stable networks that avoids this curse of dimensionality, an important open problem in the literature.

4. (Identification and Estimation) We study the empirical content of the model by characterizing the identified set based on a tractable class of pair-level network moments corresponding to the frequencies with which *dyadic outcomes* occur. The dyadic outcome of a node pair  $(i, j)$  is the tuple consisting of their observed link and network-dependent regressors that enter their joint surplus from linking. This bears some resemblance to the notion of *network type* defined by [de Paula et al. \(2015\)](#), except dyadic outcomes are defined at the level of the node pair, rather than the individual node. We find that (1) unlike moments for many-games asymptotics, computing our moments does not require the computation of the set of equilibrium networks and is therefore computationally tractable for large games, and (2) not all moments may have consistent estimators, depending on the model, which illustrates the importance of developing asymptotic theory when it comes to characterizing identified quantities in large games.

Several papers study inference in network-formation models when the econometrician observes a single network. [Chandrasekhar and Jackson \(2015\)](#) propose a new class of random graphs generated as the union of small subnetworks and develop asymptotic theory. [Boucher and Mourifié \(2013\)](#) draw on the spatial literature and provide conditions under which certain network statistics constitute a mixing random field. [Leung \(2015\)](#) studies strategic models with incomplete information. [Dzemska \(2014\)](#) and [Graham \(2014\)](#) consider dyadic link-formation models that allow for unrestricted unobserved heterogeneity. [Christakis et al. \(2010\)](#), [Hsieh and Lee \(2012\)](#), and [Mele \(2015\)](#) propose Bayesian inference procedures for dynamic models of network formation. Lastly, [de Paula et al. \(2015\)](#), [Miyauchi \(2013\)](#), and [Sheng \(2014\)](#) develop moment inequalities for network-formation games based on moments distinct from those considered in this paper.

In the next section, we state the formal assumptions of the model and then outline the intuition behind our approach in §3. We formalize these ideas in §4 and present the some of the main results in §5, namely the law of large numbers, asymptotic properties of networks generated by our model, and an algorithm for simulating counterfactual networks. In §6, we characterize the identified set and construct consistent estimators. We conduct a Monte Carlo study in §7 to study the informativeness of the derived bounds. Finally, §8 concludes.

---

<sup>4</sup>In practice, researchers generate pairwise-stable networks using myopic best-response dynamics, repeatedly iterating through each pair of nodes. This method is known to be slow for large networks and the computational complexity of the algorithm is unknown.

## 2 Setup

The econometrician observes a set of nodes or agents, the network they form, and node attributes. A model of network formation is defined by agents' preferences over networks, known up to a finite-dimensional parameter  $\theta_0$ .

**Nodes.** Let  $\mathcal{N} \subseteq \mathbb{R}^d$  be a random, locally finite set of identically distributed and almost surely unique elements.<sup>5</sup> Contrary to the usual convention of labeling nodes by natural numbers, we find it convenient to associate each node with an element of  $\mathcal{N}$ , which we term its *position*. As we discuss in §4.1, the interpretation of positions depends on the model. In some models (see Example 1 immediately below), positions are vectors of continuously distributed homophilous attributes, including, for example, a node's geographic location or income.<sup>6</sup> However, if such attributes are not available, then under some assumptions, positions can simply be labels with no intrinsic significance.

Labeling nodes by positions leads to a unified model that nests both finite and limit models, which are obtained simply by changing the stochastic process that generates node positions. In the *finite model*, for  $n \in \mathbb{N}$ , we take  $\mathcal{N} = \mathcal{N}_n$ , a set of  $n$  i.i.d. random vectors with common density  $f$  bounded above on its support. When  $n$  equals the number of nodes observed by the econometrician, this is the data-generating process for the observed network. We introduce the limit model in §4.3.

**Attributes.** For any  $i, j \in \mathcal{N}$ ,  $i$  is endowed with a vector of node-level attributes  $Z_i$ , which may include characteristics such as race or gender, and  $(i, j)$  is endowed with a vector of pair-level attributes  $\zeta_{ij}$ , which includes an idiosyncratic random shock. We assume that for any  $B \subseteq \mathbb{R}^d$ ,  $\{Z_i; i \in B\}$  and  $\{\zeta_{ij}; i, j \in B\}$  are respectively independently distributed and independent of  $\mathcal{N}$ . Note that  $\zeta_{ij}$  can depend on, for example, the distance  $\|i - j\|$ . When we discuss inference, we will assume that a subvector of  $W_{ij} = (Z_i, Z_j, \zeta_{ij})$  is unobserved by the econometrician. We let  $W = (W_{ij}; i, j \in \mathcal{N})$ .

**Network.** A network or graph is a (potentially random) function  $\Gamma : \mathcal{N} \times \mathcal{N} \rightarrow \{0, 1\}$ .<sup>7</sup> When the output is one, we say the two nodes are linked. For any pair  $i, j \in \mathcal{N}$ , we call  $\Gamma_{ij} \equiv \Gamma(i, j)$  a *potential link*. Following the usual convention, we require that  $\Gamma_{ii} = 0$  for all  $i \in \mathcal{N}$ , meaning that there are no self links.

For any  $i, j \in \mathcal{N}$  and network  $\Gamma$ , define  $S$  as the functional that maps  $(i, j, \Gamma, W, \mathcal{N})$  to a real vector, which we abbreviate as  $S_{ij}(G, W)$ , or more simply  $S_{ij}$ . An *equilibrium* network  $G$  obeys the following *pairwise-stability* condition: for every  $i, j \in \mathcal{N}$ ,

$$G_{ij} = 1 \Leftrightarrow V(S_{ij}, W_{ij}; \theta_0) > 0. \quad (1)$$

The *joint surplus function*  $V$  is known up to a finite-dimensional parameter  $\theta_0$ . We call  $S_{ij}(G, W)$  the vector of *endogenous statistics* that determine network formation,

<sup>5</sup>A set  $S \subseteq \mathbb{R}^d$  is locally finite if  $|S \cap B| < \infty$  for any bounded  $B \subseteq \mathbb{R}^d$ .

<sup>6</sup>Homophily is the pervasive phenomenon in real-world social networks that similar individuals tend to associate.

<sup>7</sup>Note that  $\Gamma$  is implicitly a function of  $\mathcal{N}$ , but we will suppress this in the notation.

due to its dependence on the endogenous state of the network  $G$ . This dependence is often referred to as *network externalities* or *strategic interactions*.

The joint surplus function is required to satisfy the following mild condition.

**Assumption 1.** *For any  $z \in \text{supp}(Z_i)$ , there exist  $z_1, z_2 \in \text{supp}(Z_i)$  and  $\zeta_1, \zeta_2 \in \text{supp}(\zeta_{ij})$  such that*

$$\max \{V(s, (z, z_1, \zeta_1); \theta_0), V(s, (z_2, z, \zeta_2); \theta_0)\} \leq 0$$

for any  $s$  in the range of  $S$ .

This is satisfied, for example, if  $\zeta_{ij}$  is unidimensional and additively separable with full support.

**Remark 1.** This setup accommodates both directed and undirected networks. In the directed case,  $V$  is the marginal utility that node  $i$  enjoys from linking with node  $j$ . When the network is undirected, we require  $V(S_{ij}, W_{ij}; \theta_0) = V(S_{ji}, W_{ji}; \theta_0)$ , and (1) corresponds to the solution concept of pairwise stability with transferable utility. Then  $V$  can be interpreted as a marginal joint surplus function. The restriction to transferable utility is solely for expositional convenience. The analysis readily extends, with minor modifications,<sup>8</sup> to the model in which  $G_{ij} = 1 \Leftrightarrow V(S_{ij}, W_{ij}; \theta_0) > 0$  and  $V(S_{ij}, W_{ij}; \theta_0) > 0$ , which corresponds to nontransferable utility. In this case, as in the directed-network setting,  $V$  is interpreted as a marginal payoff function.

The main restriction on endogenous statistics is that they satisfy a common locality restriction.

**Assumption 2** (Local Externalities). *For any  $i, j \in \mathcal{N}$  and  $G, G', W, W'$  such that  $G_{kl} = G'_{kl}$  and  $W_{kl}G_{kl} = W'_{kl}G'_{kl}$  for  $k \in \{i, j\}$  and  $l \in \mathcal{N}$ , it is the case that  $S_{ij}(G, W) = S_{ij}(G', W')$ .*

This condition restricts the dependence of the joint surplus on the state of the network, as externalities may only reach the level of indirect links. Most of the models studied in the econometric literature obey this restriction. (Christakis et al., 2010; Goldsmith-Pinkham and Imbens, 2013; Mele, 2015; Sheng, 2014). Some can allow for higher-order externalities that reach beyond a node's immediate network neighborhood (Boucher and Mourifié, 2013; de Paula et al., 2015), but most examples of interest satisfy local externalities.

---

<sup>8</sup>The definition of  $E_{ij}(r)$ , given below, and quantities that depend on it, only require slight alterations.

**Example 1.** Suppose  $G$  represents a friendship network and positions correspond to geographic locations, so  $\mathcal{N} \subseteq \mathbb{R}^2$ . Consider the model

$$V(S_{ij}, W_{ij}; \theta_0) = \theta_1 + \theta_2 \rho(i, j) + \theta_3 \max_k G_{ik} G_{jk} + \theta_4 (K_{ij} + K_{ji}) + \alpha_i + \alpha_j + \zeta_{ij},$$

where  $\theta_2 \rho(i, j)$  penalizes large distances between node positions and  $K_{ij} = \min \left\{ \sum_{k \neq j} G_{ik}, \bar{L} \right\}$ . Let  $\|\cdot\|$  be a norm on  $\mathbb{R}^2$ . Two leading examples of  $\rho$  include  $\rho(i, j) = \|i - j\|$  and

$$\rho(i, j) = \begin{cases} 0 & \text{if } \|i - j\| \leq r \\ -\infty & \text{if } \|i - j\| > r \end{cases} . \quad (2)$$

In the latter example, links only form among geographic neighbors, those located less than distance  $r$  apart from each other. In the former example, rare long-distance links may form.<sup>9</sup> In this model,  $W_{ij} = (\alpha_i, \alpha_j, \zeta_{ij})$ .

The parameter  $\theta_3$  captures *transitivity* or *clustering*, the tendency for individuals with friends in common to become friends. The importance of transitivity is widely recognized (Christakis et al., 2010; Goldsmith-Pinkham and Imbens, 2013; Jackson, 2010). The parameter  $\theta_4$  represents the importance of popularity or high degree; if  $\theta_4 > 0$ , then individuals prefer to be friends with those who have many friends, a phenomenon also known as preferential attachment (Barabási and Albert, 1999). Note that degree is truncated after some fixed number  $\bar{L}$ , which is computationally convenient for the inference procedure discussed in §6. Finally, the random effects  $\alpha_i$  and  $\alpha_j$  allow for *degree heterogeneity* (Graham, 2014), the unobserved tendency for some individuals to form more links than others.

### 3 Main Idea

The primary goal of this paper is to derive a weak law of large numbers for network moments. Such moments include average degree,  $\frac{1}{n} \sum_{i,j} G_{ij}$ , the average number of links formed by nodes in  $G$ . We define network moments and discuss additional examples in §5. Moments such as average degree can be informative for  $\theta_0$  and can be used to construct the identified set of parameters, as discussed in §6. In this section, we outline the intuition behind our approach for generating conditional independence between certain subnetworks of  $G$ , a result that forms the basis of our asymptotic theory.

The difficulty of proving a weak law for moments of equilibrium networks is that network externalities generate statistical dependence between links. Indeed, despite the fact that externalities are local in the sense of Assumption 2, the perturbation

---

<sup>9</sup>The parameter  $r$  is discussed further in §4.1. It can be point estimated, as we discuss in Remark 5.

of a single link can still “propagate” and radically change the structure of the entire network. In other words, links can be globally dependent despite the local nature of strategic interactions. This is easily understood in the context of an example. Interpret positions as geographic locations, so  $\mathcal{N} \subseteq \mathbb{R}^2$ . Let  $\Pi_{ij} = \mathbf{1}\{\|i - j\| \leq r\}$ . Viewing links as friendships between individuals, consider the following model of the joint surplus:

$$V(S_{ij}, W_{ij}; \theta_0) = \theta_1 \mathbf{1}\{G_{ab} = \Pi_{ij} \forall a = i, j, b \neq i, j\} + \theta_2 \rho(i, j) + \zeta_{ij}, \quad (3)$$

where  $\theta_1, \theta_2 > 0$  and  $\rho(i, j)$  is given by (2), meaning that  $i$  and  $j$  form a friendship only if  $\Pi_{ij} = 1$ , i.e.  $i$  and  $j$  are geographic neighbors. In this model,  $S_{ij}$  is the indicator multiplying  $\theta_1$ , which incentivizes friendships to form between those who are friends with their geographic neighbors. Let us first consider an extreme case, where  $\zeta$  is realized such that  $-\theta_1 < \zeta_{ij} < 0$  for all  $i, j \in \mathcal{N}$ .

Notice that in model (3),  $G_{ij} = 1$  only if  $\Pi_{ij} = 1$ . Hence,  $G$  is a subnetwork of the *random geometric graph* (RGG) that links  $i$  and  $j$  if and only if  $\|i - j\| \leq r$ . Suppose that the RGG is realized as in Figure 1, where a dotted line between two nodes represents a link.

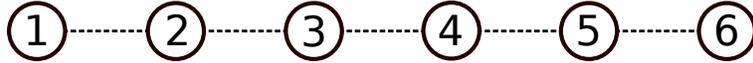


Figure 1: Random geometric graph.

First consider myopic best-response dynamics starting from the “complete” network in which  $G$  equals the RGG. Since  $\theta_1 + \zeta_{ij} > 0$ , no pair will sever their friendships, so the network is pairwise-stable. Next, consider the same dynamics but under the assumption that  $\zeta_{12} \leq -\theta_1$ , which means the pair (1, 2) does not form a friendship. This instigates (2, 3) to sever their friendship, since their joint surplus is only  $\zeta_{23} < 0$ , and so on, until the network is empty (Figure 2). Hence, the realization of  $\zeta_{12}$  in part determines the existence of links in the entire network, which is the sense in which a single link perturbation can “propagate” throughout a network.

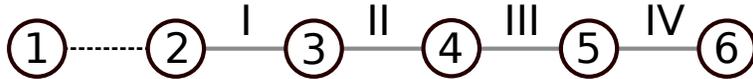


Figure 2: Roman numerals denote the order of myopic link deletions triggered by the deletion of  $G_{12}$ .

The key insight of this paper is that there often exist links that form irrespective of the state of the network, which we term *exogenously realized links*, that generate

## A WEAK LAW FOR NETWORK MOMENTS

conditional independence between sets of links. Suppose that  $\zeta_{34} > 0$ . Then regardless of  $G$ , and hence the realization of their endogenous statistics  $S_{34}$ , the pair  $(3, 4)$  will always form a friendship. Further, notice that if  $\zeta_{12} \leq -\theta_1$  in this case,  $G_{34}$  acts as a barrier that stops the propagation of externalities, so that the end result of the myopic link deletions triggered by the removal of  $G_{12}$  is Figure 3.

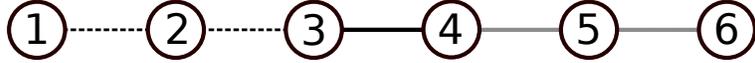


Figure 3: The black link is exogenously formed.

Because  $G_{34}$  forms exogenously, the subnetwork on  $(4, 5, 6)$  does not depend on the realization of  $\zeta_{12}$ , or for that matter, the subnetwork on  $(1, 2, 3)$ . In this sense, subnetworks on  $(1, 2, 3)$  and  $(4, 5, 6)$  are independent conditional on any path in the RGG connecting them having an exogenously realized link.

To formalize this idea, define the event that  $G_{ij}$  is exogenously realized as

$$E_{ij}(r) = \left\{ \inf_s V(s, W_{ij}; \theta_0) > 0 \cup \sup_s V(s, W_{ij}; \theta_0) \leq 0 \right\}. \quad (4)$$

Note that if the infimum of  $V$  above is positive, then  $G_{ij} = 1$  regardless of the state of the network, and we say that the link *exogenously forms*. If the supremum of  $V$  is negative, then  $G_{ij} = 0$  regardless of the network, and we say the link *exogenously fails to form*. In either case, the link  $G_{ij}$  functions as a barrier that limits cascades of link alterations.

Define  $D(r)$  as the artificial network of links that are not exogenously realized, where  $D_{ij}(r) = \Pi_{ij} \mathbf{1}\{E_{ij}^c(r)\}$  for any  $i, j \in \mathcal{N}$ . In the previous example,  $D(r)$  is generated by taking the RGG and deleting links between pairs for which the complement of  $E_{ij}(r)$  occurs. (In the next section, we will discuss a more general model of  $\Pi_{ij}$  that includes RGGs.) For instance, in Figure 3,  $D$  has two components:  $\{1, 2, 3\}$  and  $\{4, 5, 6\}$ . We will impose conditions under which

- (A)  $D(r)$  is *subcritical*, or sufficiently “fragmented” in the sense that each of its components is vanishingly small relative to  $n$ , and
- (B) components of  $D(r)$  independently form equilibrium subnetworks, conditional on attributes.

Then  $G$  is composed of a large number of conditionally independent subnetworks “stitched together” by exogenously realized links. Crucially note that this is fully compatible with the existence of a giant component in  $G$ , which will result if enough of the components of  $D$  form sufficiently connected equilibrium subnetworks, and

exogenous links form across nodes in distinct components. Idea (A) formally corresponds to requiring that  $D(r)$  *fails to percolate*, formalized in §4.4. Idea (B) is formalized in §4.2, which allows for a general class of selection mechanisms that includes the myopic best-response dynamics we used in the example.

## 4 Conditional Independence

We next outline conditions that formalize the intuition for conditional independence between subnetworks, detailed in §3.

### 4.1 Opportunity Graph

We assume the joint surplus encodes an exogenous *opportunity graph*  $\Pi$  that describes the pairs of nodes that have opportunities to form links. A special case of  $\Pi$  is the random geometric graph discussed in §3. We will first formally define such graphs and then give additional examples below.

**Assumption 3** (Opportunity Graph). *There exists a network  $\Pi$ , which we call an opportunity graph such that (a) for any  $i, j \in \mathcal{N}$ ,*

$$\Pi_{ij} = \mathbf{1} \{ \mu(\|i - j\|, W_{ij}; r) \geq 0 \}$$

*for some function  $\mu$ , norm  $\|\cdot\|$  on  $\mathbb{R}^d$ , and  $r \equiv r(\mathcal{N}) \in \mathbb{R}^+$ , and (b)*

$$\mathbf{P}(V(S_{ij}, W_{ij}; \theta_0) > 0 \mid \Pi_{ij} = 0) = 0. \tag{5}$$

Equation (5) states that a pair of nodes may only link if they have the opportunity to do so, as dictated by  $\Pi$ . Hence, under this assumption, the equilibrium network  $G$  implicitly depends on  $r$ . The choice of the parameter  $r$  will depend on the stochastic process generating  $\mathcal{N}$ , as we discuss in §4.3. Opportunity graphs play an important role in our analysis because the asymptotic properties of  $G$  are in part determined by  $\Pi$ . In particular,  $G$  is sparse if  $\Pi$  is sparse, and sparsity of  $\Pi$  will be needed to establish a law of large numbers. §4.3 and §4.4 will impose additional restrictions on  $\Pi$ , including sparsity.

**Example 2** (Random Geometric Graph). The random geometric graph (RGG) model is obtained by setting  $\mu(\|i - j\|, W_{ij}; r) = 1 - r^{-1}\|i - j\|$ , so  $\Pi_{ij} = \mathbf{1}\{\|i - j\| \leq r\}$ . See Penrose (2003) for a comprehensive survey of the RGG literature. The RGG is the opportunity graph in model (3) and Example 1 when  $\rho$  is given by (2). In the context of social networks, node positions in this model represent continuous homophilous attributes. For instance, if position corresponds to geographic location, then  $\|i - j\|$  is geographic distance, and the model states that nodes only link with those in a

fixed geographic radius, which captures geographic homophily. Position may include other continuous attributes, such as income, but the leading case of interest is when position corresponds to geographic location, since geographic homophily is widespread in social networks. In development contexts or student friendship networks, agents are strongly constrained in terms of geographic mobility. Physician referral networks are another example, since doctors are unlikely to refer patients to specialists a substantial distance away. Venture capital coinvestment networks are also geographically local in nature (Sorenson and Stuart, 2001; Uetake, 2012).

Another case of interest is when node positions are latent. A well-known literature in statistics studies such “latent-space” models (Hoff et al., 2002), interpreting  $\mathcal{N}$  as positions in “social space,” so that only socially close nodes form links. These models have been applied in political science and sociology (Hoff and Ward, 2004; Treier and Jackman, 2008).

Below we will require  $r \rightarrow 0$  at a particular rate as  $n \rightarrow \infty$  to ensure sparsity; this is a typical asymptotic regime of interest when studying RGGs. This can be interpreted as nodes become increasingly selective about their partners the larger the pool of available partners. The selectivity story is our rationalization of network sparsity, the fact that the typical node does not link with the vast majority of other nodes in the network. An alternative interpretation is attained by rescaling positions and  $r$  by  $r^{-1}$ , so that  $\Pi_{ij} = \mathbf{1}\{\|r^{-1}i - r^{-1}j\| \leq 1\}$ . This is a version of “increasing domain” asymptotics in the spatial literature, where homophilic preferences are now fixed in  $n$ , but the set of node positions is  $r^{-1}\mathcal{N}$ , so nodes become increasingly *diverse*, or spread out, as  $n$  grows. In contrast, under the selectivity story, the space of node positions is fixed, as under “infill” asymptotics, but preferences change with  $n$ . Both interpretations achieve the same effect of ensuring network sparsity.

**Example 3** (Random Connection Model). The random connection model (RCM) generalizes the RGG model to allow for rare “long-distance” links. The main reference for this model is Meester and Roy (1996), who discuss the percolative properties of RCMs. Let  $\mu(\|i - j\|, W_{ij}; r) = \mu^*(r^{-1}\|i - j\|, W_{ij})$ . In Example 1, when  $\rho(i, j) = \|i - j\|$ , we can define

$$\mu^*(r^{-1}\|i - j\|, \eta_{ij}) = \sup_s V(s, W_{ij}; \theta_0) = \theta_1 + \theta_2 \rho(i, j) + \theta_3 + \theta_4 \bar{L} + \alpha_i + \alpha_j + \zeta_{ij},$$

where  $\theta_2 = -r^{-1}$ . In this simple linear model, if  $\zeta_{ij}$  has full support, then links may occur even if  $\|i - j\|$  is large. Thus, while RGGs assume a “hard threshold,” where links never occur if  $\|i - j\|$  exceeds a certain threshold, the random connection model allows for “soft” thresholds, where the linking probability tends to zero as distance  $\|i - j\|$  tends to infinity. As with RGGs,  $r$  will tend to zero at a certain rate to ensure sparsity.

**Example 4** (Erdos-Renyi Graph). A general way of imposing Assumption 3 is to

assume that

$$G_{ij} = 1 \Leftrightarrow \tilde{V}(S_{ij}, W_{ij}; \theta_0) \Pi_{ij} > 0, \quad (6)$$

where  $\tilde{V}$  is the joint surplus, and  $\Pi$  might be interpreted as an exogenous “meeting process.” While the general model of  $\Pi$  allows for node-level heterogeneity, a simple example of a meeting process is independent random meetings, an assumption often made in the theoretical literature (Jackson and Watts, 2002). Then  $\Pi$  is generated by a classical Erdős-Rényi random graph, which we refer to as the ER model.

In this model,  $\mu(\|i - j\|, W_{ij}; r) = p(r) \mathbf{1}\{i, j \in [0, 1]\} - \eta_{ij}$ , where  $\eta_{ij} \stackrel{iid}{\sim} U[0, 1]$  is a subvector of  $W_{ij}$ . Then pairs of nodes meet independently with probability  $p(r)$ . Sparsity of  $\Pi$  is achieved when  $p(r)$  is of order  $n^{-1}$ , so the expected number of meetings for any given node is a constant. Unlike the previous examples, this model of  $\Pi$  does not require any form of homophily, and node positions are merely labels rather than homophilous attributes.

**Example 5** (Stochastic Block Model). The stochastic block model (SBM) generalizes ER models by assuming the existence of  $K < \infty$  “groups,” where the probability of a link opportunity depends only on the groups of the ego and alter. In contrast to RGGs, this model allows for homophily in discrete attributes. Inference for SBMs is a rapidly growing area of research in the statistics literature (e.g. Bickel and Chen, 2009; Bickel et al., 2011). See Bollobás et al. (2007) for a comprehensive study of their percolative properties.

Suppose  $\alpha_i$ , a scalar subvector of  $Z_i$ , represents the group membership of node  $i$ , with memberships distributed i.i.d., as are the disturbances  $\eta_{ij}$ , with  $\eta_{ij} \perp (\alpha_i, \alpha_j)$ . As in ER models, node positions are merely labels and do not represent characteristics. The stochastic block model is

$$\mu(\|i - j\|, W_{ij}; r) = \rho(\alpha_i, \alpha_j, r) \mathbf{1}\{i, j \in [0, 1]\} - \eta_{ij},$$

where the codomain of  $\rho$  is  $[0, 1]$ . Then conditional on  $\alpha_i$  and  $\alpha_j$ ,  $\Pi(i, j)$  is independently distributed with link formation probability  $\rho(\alpha_i, \alpha_j, r)$ . As with ER models, sparsity of  $\Pi$  will require  $\rho(\alpha_i, \alpha_j)$  to be of order  $n^{-1}$  with probability one.

## 4.2 Equilibrium Selection

Model (1) does not fully define a likelihood for  $G$  because conditional on  $W$ , there may be multiple equilibrium networks that satisfy (1) (see e.g. Sheng, 2014, for examples). In order to complete the model, we introduce a selection mechanism, which maps the set of possible equilibria to a single network. In addition, we will impose a key restriction on the selection mechanism that generates conditional independence across certain subnetworks; see idea (B).

Define  $\mathcal{G}_{\theta, r}(W, \mathcal{N})$  to be the correspondence that maps  $(W, \mathcal{N}, \theta, r)$  to the set of equilibrium networks on  $\mathcal{N}$ . We introduce a random vector  $\nu$ , which we interpret

## A WEAK LAW FOR NETWORK MOMENTS

as a *public signal* that, along with  $W$  and  $\mathcal{N}$ , nodes utilize to “coordinate” on the formation of an equilibrium network. As a simple example, suppose for some given  $W$ ,  $\mathcal{N}$ , and  $\theta$  there are two possible equilibrium networks, and nodes coordinate on a particular network by flipping a coin. The outcome of the flip is represented by  $\nu$ . Thus, unlike  $W$  and  $\mathcal{N}$ ,  $\nu$  is a signal that does not directly enter  $V$  but still affects link formation through equilibrium selection. With this notation, we can define a *selection mechanism* as a function  $\lambda : (W, \mathcal{N}, \nu; \theta, r) \mapsto G \in \mathcal{G}_{\theta, r}(W, \mathcal{N})$ . Our first assumption just defines the basic requirement that some selection mechanism rationalizes the data.

**Assumption 4** (Selection Mechanism). *For any  $\mathcal{N}$  and  $W$ , the following hold with probability one.*

- (a) (Coherence)  $|\mathcal{G}_{\theta_0, r}(W, \mathcal{N})| \geq 1$ .
- (b) (Rationalizability) *There exists a random vector  $\nu \perp (W, \mathcal{N})$  and selection mechanism  $\lambda$  such that for any equilibrium network  $G$ ,*

$$G = \lambda_{\theta_0, r}(W, \mathcal{N}, \nu).$$

Part (a) states that an equilibrium network exists. Without this assumption, the econometrician must take a stance on the realization of  $G$  when there are no equilibrium networks. When  $\mathcal{N}$  is finite, there are sufficient conditions in the literature that guarantee equilibrium existence (Sheng, 2014). In the limit model,  $\mathcal{N}$  is infinite, but we show that under certain conditions, the global network consists of “stitched together” equilibrium subnetworks that are almost surely finite. Hence, the same existence results for finite  $\mathcal{N}$  may be employed. This is discussed further below.

**Remark 2.** Selection mechanisms are more commonly represented as conditional distributions  $\sigma$  on  $\mathcal{G}_{\theta, r}(W, \mathcal{N})$ . Our definition is equivalent. To see this, fix  $W$  and  $\mathcal{N}$ , and for simplicity suppose that  $\mathcal{G}_{\theta, r}(W, \mathcal{N}) = \{G_1, G_2\}$ . Let  $\sigma_{\theta, r}(G | W, \mathcal{N})$  place probability  $p \equiv p_{\theta}(W, \mathcal{N})$  on  $G_1$ . Now let  $\nu \sim U[0, 1]$ , independent of  $W$  and  $\mathcal{N}$  and define  $\lambda_{\theta, r}(W, \mathcal{N}, \nu)$  to equal  $G_1$  if  $\nu \in [0, p]$  and  $G_2$  otherwise. Then clearly  $\sigma_{\theta, r}$  is the distribution of  $\lambda_{\theta, r}$ .

We next impose a restriction on the selection mechanism that formalizes idea (B). In example (3), the endogenous statistics do not depend on attributes  $W$ . Assumption 2 allows for such dependence, which leads to an additional source of correlation. In Figure 3, for example, the surpluses enjoyed by nodes (1, 2, 3) may depend on  $Z_4$ , as may the surpluses enjoyed by nodes (4, 5, 6). This motivates the following notion of “augmented” components of  $D(r)$ , which we will assume form subnetworks in  $G$  independently.

Recall that  $D(r)$  is defined such that  $D_{ij}(r) = \Pi_{ij} \mathbf{1}\{E_{ij}^c(r)\}$ , where  $E_{ij}(r)$  is defined in (4). For each  $i \in \mathcal{N}$ , let  $C_i(r) \equiv C(i, \mathcal{N}, D, r)$  be the set of nodes in  $i$ 's component under  $D(r)$  and

$$C_i^+(r) \equiv C^+(i, \mathcal{N}, D, r) = C_i(r) \cup \{j \in \mathcal{N} : \exists k \in C_i(r) \text{ such that } \Pi_{ij} = 1\}.$$

We refer to  $C_i^+(r)$  as  $i$ 's *augmented component*. Define  $\mathcal{C}^+(D, \mathcal{N}, r) = \{C_i^+(r) : i \in \mathcal{N}\}$ , the set of augmented components. To understand augmented components, consider again the random geometric graph in Figure 1 and model (3). Suppose that  $\zeta_{23}, \zeta_{34}, \zeta_{45} > 0$ , so that the associated node pairs exogenously form links, while  $\zeta_{ij} \in (-\theta, 0)$  for all other pairs. Figure 4 depicts the resulting equilibrium network. The components of  $D$  in the model are  $\{1, 2\}, \{3\}, \{4\}, \{5, 6\}$ , and there are four augmented components:  $\{1, 2, 3\}, \{2, 3, 4\}, \{3, 4, 5\}, \{4, 5, 6\}$ . Notice that while the set of components  $\{C_i(D, r) : i \in \mathcal{N}\}$  constitutes a partition of  $\mathcal{N}$ , the set of augmented components  $\mathcal{C}^+(D, \mathcal{N}, r)$  may not.

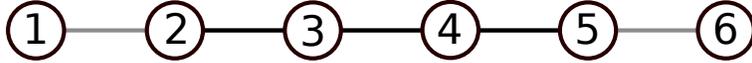


Figure 4: Black links represent exogenously formed links.

A few last pieces of notation are required. For any  $A \subseteq \mathcal{N}$ , let  $W_A = \{W_{ij} : i, j \in A\}$ . For  $C \in \mathcal{C}^+(D, \mathcal{N}, r)$ , let  $\lambda_{\theta, r}(W, \mathcal{N}, \nu)|_C$  be the restriction of the range of  $\lambda_{\theta, r}$  to equilibrium subnetworks on  $C$ , i.e.  $\mathcal{G}_{\theta, r}(W_C, C)$ . This is well defined, since by Assumption 2, the joint surplus of a pair of nodes in the same augmented component does not depend on attributes or links formed by nodes outside of this augmented component. For instance, in Figure 4, we can consider the set of equilibrium subnetworks formed on  $\{4, 5, 6\}$  in isolation because attributes and links in the rest of the network do not affect this set.

**Assumption 5** (No Coordination). *Let  $\mathcal{C}^+(D, \mathcal{N}, r) = \{C_1, C_2, \dots\}$ .<sup>10</sup> With probability one, there exist  $\nu_1, \nu_2, \dots$  independently distributed and independent of  $W$  and  $\mathcal{N}$  such that for any  $k \in \mathbb{N}$ ,*

$$\lambda_{\theta, r}(W, \mathcal{N}, \nu)|_{C_k} = \lambda_{\theta, r}(W_{C_k}, C_k, \nu_k).$$

Assumption 5 is satisfied when  $\lambda$  is *degenerate*, meaning it does not depend on  $\nu$ . It is also satisfied if the network is formed via myopic best-response dynamics, which in finite models have been shown to converge to a pairwise-stable equilibrium in finite time (Jackson and Watts, 2002).

<sup>10</sup>The set of components is countable because  $\mathcal{N}$  is countable.

This assumption states that augmented components independently “coordinate” on their respective equilibrium subnetworks. Again, the idea is that best-response dynamics do not percolate beyond a node’s augmented component due to the existence of exogenously realized links. To see what this assumption rules out, consider Figure 3, and suppose that for some given  $W$  and  $\theta$ , there exist two equilibrium subnetworks on  $(1, 2, 3)$  and  $(4, 5, 6)$ , respectively. An example of coordination is if the two triplets flip a coin, and both collectively decide to play a particular equilibrium if the coin flip is heads and to play the other equilibrium if the flip is tails. Assumption 5 requires instead that the two triplets independently flip coins. This is sensible because the joint surplus (3) enjoyed by nodes  $(1, 2, 3)$ , for any equilibrium subnetwork, does not depend on the network played by nodes  $(4, 5, 6)$  by Assumption 2. Hence, the triplets have no incentive to coordinate. Notice that a similar assumption is required even in the case when the econometrician observes a large cross section of networks; in order for two network observations to be independent, there must not be coordination between the two sets of nodes in question.

### 4.3 Sparsity

We next impose a rate restriction on  $r(\mathcal{N}_n)$  that ensures asymptotic sparsity of the opportunity graph. By Assumption 3, this implies sparsity of  $G$ . As we will see, a sparse opportunity graph is important for ensuring “weak” conditional dependence between node-level functions of links that define network moments (e.g. node degrees).

**Assumption 6** (Sparsity).

(a)  $r(\mathcal{N}_n) \equiv r_n$  satisfies  $nr_n^d \rightarrow \kappa < \infty$  as  $n \rightarrow \infty$ .

(b) For  $p_{ij}(r) = \mathbf{E}[\Pi_{ij}(r) | i, j]$ ,

$$n \int \int p_{ij}(r_n) f(i) f(j) di dj \rightarrow \kappa \int \int p_{ij}(1) f(i)^2 di dj < \infty. \quad (7)$$

The left-hand side of (7) is the expected degree of the opportunity graph in the finite model. To interpret the right-hand side, for each  $i$  in the support of  $f$ , we define the following *limit model*:

- The set of nodes is generated by a Poisson process on  $\mathbb{R}^d$  with intensity  $\kappa f(i)$ , i.e.  $\mathcal{N} = \mathcal{P}_{\kappa f(i)}$ .
- The associated opportunity graph parameter  $r(\mathcal{P}_{\kappa f(i)})$  equals unity.

Then for any fixed  $i$ ,  $\int \kappa p_{ij}(1) f(i) dj$  is evidently the expected degree of  $\Pi$  when  $\mathcal{N} = \mathcal{P}_{\kappa f(i)}$ . The right-hand side of (7) takes the expectation of this quantity with respect to  $i$ .

In general, we can establish (7) using Lemma 3 in the appendix, provided  $\mu$  and  $W$  satisfy certain stationarity conditions detailed in §A.1. We next verify Assumption 6 for several examples of opportunity graphs, each of which satisfies these invariance conditions.

**Example 6** (Erdos-Renyi Graph). In this model, we let  $f$  be the uniform density on  $[0, 1]$ , and, following the setup in Example 4, assume that  $i$  and  $j$  meet independently with probability  $p_{ij}(r) = r_n^d \mathbf{1}\{i, j \in [0, 1]\}$ , where  $nr_n^d = \kappa$  for all  $n$ . Then the expected degree of a node is  $\kappa$ . To check that (7) holds, simply note that the left-hand side is the expected degree of a network with  $n$  nodes, which is  $\kappa$ , and the right-hand side is  $\kappa \int \int \mathbf{1}\{i, j \in [0, 1]\} f(i)^2 di dj$ , which also equals  $\kappa$ , since  $f$  is the uniform density. Note that the limit model here accords with the usual limit model for Erdős-Rényi graphs in which the number of connections formed by a given node follows a Poisson( $\kappa$ ) distribution.

**Example 7** (Stochastic Block Model). For  $\rho$  defined in Example 5, we take

$$\rho(\alpha_i, \alpha_j, r_n) = \tilde{\rho}(\alpha_i, \alpha_j) \mathbf{1}\{i, j \in [0, 1]\} r_n^d$$

for a function  $\tilde{\rho}$  with codomain  $\mathbb{R}^+$ . We set  $r_n = n^{-d}$  and assume positions are drawn uniformly from  $[0, 1]$  and  $\alpha_i$  is distributed i.i.d. with  $K < \infty$  support points. Since  $\Pi_{ij} \sim \text{Ber}(\rho(\alpha_i, \alpha_j)n^{-1})$ , the left-hand side of (7) is  $\mathbf{E}[\tilde{\rho}(\alpha_i, \alpha_j)]$ , which is equivalent to the right-hand side since  $f$  is the uniform density.

**Example 8** (Random Geometric Graph). The requirement  $nr_n^d \rightarrow \kappa$  implies  $r_n \rightarrow 0$ . We claim that  $nr_n^d \rightarrow \kappa$  suffices for (7). To see this, define  $B(i, r) = \{j \in \mathcal{N} : \|i - j\| \leq r\}$ , and  $\text{Vol}(B(0, 1))$  as the volume of the unit ball in  $\mathbb{R}^d$  centered at zero with respect to  $\|\cdot\|$ . Then the left-hand side of (7) is equal to the expected degree of  $\Pi$  under  $\mathcal{N}_n$ , namely

$$\begin{aligned} \mathbf{E} \left[ \frac{1}{n} \sum_{i,j} \mathbf{1}\{\|i - j\| \leq r_n\} \right] &= n \mathbf{E}[\mathbf{P}(\|i - j\| \leq r_n | i)] = \int n \int_{j \in B(i, r_n)} f(j) dj f(i) di \\ &= \int nr_n^d r_n^{-d} \underbrace{\int_{j \in B(i, r_n)} [f(j) - f(i)] dj}_{o(1)} f(i) di + \int n \int_{B(i, r_n)} f(i) dj f(i) di \\ &= \int nr_n^d \text{Vol}(B(0, 1)) f(i)^2 di + o(1), \quad (8) \end{aligned}$$

where  $o(1)$  term appears due to the Lebesgue density theorem. Now, the right-hand side of (7) is simply

$$\kappa \int \int \mathbf{1}\{\|i - j\| \leq 1\} f(i)^2 di dj = \kappa \int \text{Vol}(B(0, 1)) f(i)^2 di,$$

which is the limit of (8) when  $nr_n^d \rightarrow \kappa$ , as desired.

**Example 9** (Random Connection Model). In this model, node positions are distributed according to any arbitrary density bounded above. Here

$$p_{ij}(r_n) = \mathbf{P}(\mu^*(r_n^{-1}\|i-j\|, W_{ij}) \geq 0 \mid r_n^{-1}\|i-j\|) \equiv F(r_n^{-1}\|i-j\|).$$

Proving (7) is more technical for a model as general as this, but convergence will follow from Lemma 3 in the appendix under the asymptotic sparsity condition

$$\kappa \int \int F(\|i-j\|) f(i)^2 di dj < \infty. \tag{9}$$

If  $f(\cdot)$  is uniform, then provided  $\kappa$  is finite, this is equivalent to  $\int_{x \geq 0} F(x) dx < \infty$ . If  $\mu^*$  is linear in its arguments, it is clearly necessary for the tails of  $W_{ij}$  to be sufficiently thin. For instance, if  $W_{ij}$  is standard normal, then this integral is in fact finite (see Example 11 below for further discussion).

#### 4.4 Subcriticality

Our last assumption formalizes idea (A). It implies that the expected degree of  $D$  under the limit models must be less than one, which is sufficient for subcriticality of  $D$ .

Define  $\bar{f} = \text{ess sup}_x f(x)$  and

$$\gamma_r(i, Z_i) = \kappa f(i) \int \mathbf{P}(E_{ij}^c(r) \cap \{\Pi_{ij}(r) = 1\} \mid i, j, Z_i) dj.$$

**Assumption 7.**  $\kappa \bar{f} \|\gamma_1(i, Z_i)\|_2 < 1$ , where  $\|\cdot\|_2$  is the  $L_2$  norm, taken with respect to the random vector  $(i, Z_i)$  for  $i \sim f(\cdot)$ .

**Proposition 1** (Subcriticality). *Under Assumptions 6 and 7, for any fixed  $i, j$  in the support of  $f(\cdot)$ , with probability one,  $|C(j, \mathcal{P}_{\kappa f(i)}, D, 1)| < \infty$ .*

The result states that the largest component of  $D$  in any limit model is finite with probability one. This formalizes idea (A) that the artificial network  $D$  should be composed of many components, vanishingly small relative to the network size, for enough links in  $G$  to be conditionally independent.

Assumption 7 implies that the expected degree of  $D$  averaged over the limit models is less than one. For this to hold, either the expected degree of  $\Pi$  must be sufficiently small, or  $E_{ij}(1)$  has to hold with high enough probability, meaning that enough links must be exogenously realized. This is illustrated in the next example.

**Example 10** (Random Geometric Graph). Suppose  $d = 2$ ,  $\Pi_{ij}(1) = \mathbf{1}\{\|i - j\| \leq 1\}$  (random geometric graph), and  $f$  is the uniform density on  $[0, 1]^2$ . For simplicity, assume that  $\{\zeta_{ij}; i, j \in B\}$  is identically distributed for any  $B \subseteq \mathbb{R}^d$ . Consider the model in Example 1 with  $\rho$  defined in (2) and  $\theta_3, \theta_4 \geq 0$ . Define

$$\gamma = \|\mathbf{P}(-\theta_3 - \theta_4 \bar{L} < \theta_1 + \alpha_i + \alpha_j + \zeta_{ij} \leq 0 \mid \alpha_i)\|_2,$$

Then

$$\|\gamma_1(i, Z_i)\|_2 \leq \gamma \operatorname{ess\,sup}_i \kappa f(i) \int \mathbf{1}\{\|i - j\| \leq 1\} dj = \gamma \kappa \pi.$$

Hence, Assumption 7 is satisfied if  $\gamma < (\kappa \pi)^{-1}$ . There are two ways this condition can be satisfied:

- Case  $\kappa \pi < 1$ : The condition is satisfied for *any*  $\gamma$ , which means no further restrictions on the model are required. However,  $\kappa \pi$  is the limiting expected degree of  $\Pi(1)$ , and if this falls below one, then it is well known that  $\Pi(1)$  does not percolate (see §5.2), and therefore  $G$  does not percolate.
- Case  $\kappa \pi \geq 1$ : The condition is satisfied if  $\gamma$  is sufficiently small. This occurs if  $\theta_1 + \alpha_i + \alpha_j + \zeta_{ij}$  is sufficiently large with high probability in the sense of either exceeding zero or falling below  $-\theta_3 - \theta_4 \bar{L}$ . In other words,  $W_{ij}$  must be sufficiently positive or negative frequently enough to ensure that enough links are realized exogenously. As we discuss later, this can be compatible with percolation of  $G$ .

**Example 11** (Random Connections Model). Consider Example 1 with  $\theta_3, \theta_4 \geq 0$ ,  $\rho(i, j) = \|i - j\|$ , and positions drawn uniformly from some bounded region. Suppose for any  $i, j \in \mathbb{R}^d$ ,  $\zeta_{ij} \sim N(0, 1)$  and  $\alpha_i \sim N(0, \sigma^2)$  with  $(\alpha_i, \alpha_j) \perp \zeta_{ij}$ . Following Example 3, let

$$\begin{aligned} \Pi_{ij}(r) &= \mathbf{1}\{\theta_1 + \theta_2 \|i - j\| + \theta_3 + \theta_4 \bar{L} + \alpha_i + \alpha_j + \zeta_{ij} > 0\}, \\ D_{ij}(r) &= \mathbf{1}\{\theta_1 + \theta_2 \|i - j\| + \alpha_i + \alpha_j + \zeta_{ij} \leq 0\} \Pi_{ij}(r). \end{aligned}$$

(Recall  $\theta_2 = -r^{-1}$  in this model.)

*Sparsity.* We first check (9). Let  $\tilde{\theta} = \theta(1 + 2\sigma^2)^{-1/2}$ . In this model,

$$F(\|i - j\|) = \Phi\left(\tilde{\theta}_1 + \tilde{\theta}_2 \|i - j\| + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L}\right),$$

where  $\theta_2 = -1$  and  $\Phi(\cdot)$  is the CDF of the standard normal distribution. As noted in Example 9, since  $f(\cdot)$  is the uniform density, (9) is equivalent to  $\kappa \int_{x \geq 0} F(x) dx < \infty$ . Some calculus shows that

$$\int_0^\infty \Phi(a + bx) dx = \lim_{x \rightarrow \infty} \frac{1}{b} ((a + bx)\Phi(a + bx) + \phi(a + bx)) - \frac{1}{b} (a\Phi(a) + \phi(a)),$$

## A WEAK LAW FOR NETWORK MOMENTS

where  $\phi(\cdot)$  is the density of the standard normal distribution. If  $b < 0$ , then the limit on the right-hand side equals zero. Therefore, setting  $a = \tilde{\theta}_1 + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L}$ , and  $b = \tilde{\theta}_2$ ,

$$\begin{aligned} \int_{x \geq 0} F(x) dx &= \int_0^\infty \Phi\left(\tilde{\theta}_1 + \tilde{\theta}_2 x + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L}\right) dx \\ &= -\frac{1}{\tilde{\theta}_2} \left( \left( \tilde{\theta}_1 + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L} \right) \Phi\left(\tilde{\theta}_1 + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L}\right) + \phi\left(\tilde{\theta}_1 + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L}\right) \right), \end{aligned}$$

which is finite as desired.

*Subcriticality.* In this model,  $Z_i = \alpha_i$  and

$$\begin{aligned} \gamma_1(i, Z_i) &= \kappa f(i) \int \mathbf{P}\left(\theta_1 + \|i - j\| + \alpha_i + \alpha_j + \zeta_{ij} \leq 0 \cap \Pi_{ij}(1) = 1 \mid i, j, \alpha_i\right) dx \\ &= \kappa \int_{x \geq 0} \mathbf{P}\left(\theta_1 + x + \alpha_i < \alpha_j + \zeta_{ij} \leq \theta_1 + x + \theta_3 + \theta_4 \bar{L} + \alpha_i \mid x, \alpha_i\right) dx \\ &= \kappa \left[ \int_{x \geq 0} \Phi\left(\tilde{\theta}_1 + \tilde{\theta}_2 x + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L} + \alpha_i\right) dx - \int_{x \geq 0} \Phi\left(\tilde{\theta}_1 + \tilde{\theta}_2 x + \alpha_i\right) dx \right] \\ &= \frac{\kappa}{\tilde{\theta}_2} \left[ \left( \tilde{\theta}_1 + \alpha_i \right) \Phi\left(\tilde{\theta}_1 + \alpha_i\right) + \phi\left(\tilde{\theta}_1 + \alpha_i\right) \right. \\ &\quad \left. - \left( \tilde{\theta}_1 + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L} + \alpha_i \right) \Phi\left(\tilde{\theta}_1 + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L} + \alpha_i\right) - \phi\left(\tilde{\theta}_1 + \tilde{\theta}_3 + \tilde{\theta}_4 \bar{L} + \alpha_i\right) \right], \end{aligned}$$

where  $\tilde{\theta} = \theta(1 + \sigma^2)^{-1/2}$  and  $\theta_2 = -1$ . Clearly, the last line would be zero, and therefore Assumption 7 would be satisfied, if  $\theta_3 = \theta_4 = 0$ . Hence, for  $\|\gamma_1(i, Z_i)\| < 1$  to hold, the externality parameters must not be too large relative to  $1 + \sigma^2$ . If  $\kappa$  is small, then clearly these parameters may be larger. On the other hand, if  $\kappa$  is too small, then  $\Pi$  will not percolate.

In §5.2, we derive necessary and sufficient conditions for  $G$  to percolate for the models of  $\Pi$  presented in §4.1. We relate these conditions to Assumption 7.

**Remark 3** (Equilibrium Existence). Consider the limit model in which  $\mathcal{N} = \mathcal{P}_\tau$  for some  $\tau \in (0, \infty)$ . Under Assumption 5, the limit model generates networks by independently selecting equilibrium subnetworks on augmented components of  $D$ , which by Proposition 1 are almost-surely finite. These components are then connected by exogenously realized links. It therefore follows that conditions for equilibrium existence in the finite model guarantee existence in the limit model. Finiteness of equilibrium subnetworks on augmented components on  $D$  in the limit model plays an important role in the proof of the main theorem.

We sketch a proof of Proposition 7 for the simple case in which  $W_{ij} = \zeta_{ij}$  and  $\{\zeta_{ij}; i, j \in B\}$  is identically distributed for any  $B \subseteq \mathcal{N}$ . Under these conditions, for any  $k$  in the support of  $f$ , the events  $\{E_{ij}^c(1); i, j \in \mathcal{P}_{\kappa f(k)}\}$  are independent, and

the result can be obtained using a simple branching process argument. Begin at an arbitrary node, and branch out to its network neighbors in  $D$  (“offspring”), recording the number of such offspring  $O_1$ . Then branch out to all of their neighbors (potentially including the initial node of the process), recording the number of such neighbors  $O_2$ . The key observation is that if  $\sum_{m=1}^{\infty} O_m$  is finite, then the branching process dies out, and the initial node’s component is therefore finite, proving part (a) of the theorem. Now, notice that by construction of  $O_m$ ,  $\mathbf{E}[O_m] = \mathbf{E}[O_1]^m$ . Further, Assumption 7 implies that the expected number of offspring  $\mathbf{E}[O_1^m]$  is less than one. Therefore,  $\mathbf{E}[\sum_{m=1}^{\infty} O_m] = \frac{1}{1-\mathbf{E}[O_1]} < \infty$ . It follows that the branching process dies out with probability one.

## 5 Main Results

Our main result is a weak law for network moments that explicitly characterizes their limits. We first define these moments.

**Definition 1.** Let  $\psi$  be a functional that maps  $(i, \mathcal{N}, G, W, r)$  to a real number.<sup>11</sup> If  $i \notin \mathcal{N}$ , we abbreviate  $\psi(i, \mathcal{N}, G, W, r) \equiv \psi(i, \mathcal{N} \cup \{i\}, G, W, r)$ . We require  $\psi$  to satisfy the following *locality* restriction:

For any  $i \in \mathcal{N}$  and  $G, G', W, W', \Pi, \Pi'$  such that  $\Pi_{ij} = \Pi'_{ij}$  for all  $j \in \mathcal{N}$  and  $G_{jk} = G'_{jk}$  and  $W_{jk}G_{jk} = W'_{jk}G'_{jk}$  for all  $j, k \in \mathcal{N}$  such that  $\Pi_{ij} = 1$ , it is the case that  $\psi(i, \mathcal{N}, G, W, r) = \psi(i, \mathcal{N}, G', W', r)$ .

We call any such  $\psi$  a *node statistic*. A *network moment* is an average of node statistics  $\frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} \psi(i, \mathcal{N}, G, W, r)$  on finite subsets  $\mathcal{N}$  of  $\mathbb{R}^d$ .

The locality restriction simply states that node  $i$ ’s node statistic only depends on the network, attributes, and opportunity graph through the direct links in  $G$  of nodes  $j$  who are linked with  $i$  in the opportunity graph. A special case of this is if  $\psi_i(\mathcal{N}, r)$  only depends on  $i$ ’s direct network neighbors in  $G$  and the neighbors of her neighbors. This is similar to the local externalities restriction on  $S_{ij}$ .

**Example 12.** The degree of node  $i$  with respect to  $G$  is  $\sum_j G_{ij}$ , whereas its degree with respect to  $\Pi$  is  $\sum_j \Pi_{ij}$ . The latter is clearly a node statistic. By assumption,  $G_{ij} = 1$  only if  $\Pi_{ij} = 1$ , so  $\sum_j G_{ij}$  is therefore also a node statistic. Average degree  $\frac{1}{n} \sum_{i \neq j} G_{ij}$  is equivalent to link frequency scaled up by  $n$ . This scaling is necessary to obtain a non-degenerate limit, since link frequency tends to zero as  $n \rightarrow \infty$  by sparsity. By analogy to standard discrete-choice models, it is intuitive that its expectation, the probability of link formation, should be informative for  $\theta_0$ .

<sup>11</sup>Note that  $G$  (and possibly  $W$ ) depends on  $r$ . The last argument of  $\psi$  supplies the value of  $r$ .

**Example 13.** The *individual clustering* for a node  $i$  is

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

This is the proportion of three-node subnetworks in which  $i$  is linked to two nodes  $j$  and  $k$  that are transitive, meaning that  $j$  and  $k$  are also linked. We define  $Cl_i(G) \equiv 0$  if  $i$  has at most one link. By the same reasoning for average degree,  $Cl_i(G)$  is a valid node statistic. The *average clustering coefficient* of  $G$  is  $\frac{1}{n} \sum_{i=1}^n Cl_i(G)$ , a well-known measure of transitivity in network science. Thus, its population analog is likely informative for  $\theta_3$  in Example 1.

**Example 14.** de Paula et al. (2015) define a node  $i$ 's *network type* as the local subnetwork on  $i$  (including both links and attributes) up to path distance  $D$  away.<sup>12</sup> In order for a network type to be a valid node statistic, the locality restriction on  $\psi$  requires  $D = 2$ . de Paula et al. show that an identified set for a class of network-formation games can be constructed using network type shares, defined as the proportion of nodes of a given network type.<sup>13</sup>

Further examples of network moments will be given in §6 when we define our construction of the identified set. In particular, we will consider moments that correspond to the frequencies with which the “dyadic outcome”  $(G_{ij}, S_{ij})$  takes on a particular realization.

## 5.1 Weak Law

For  $a \in \mathbb{R}$  and  $b \in \mathbb{R}^d$ , let  $a(\mathcal{N} - b) = \{a(i - b) : i \in \mathcal{N}\}$ . We can now state the main theorem.

**Theorem 1** (Weak Law). *Suppose that  $\psi$  is uniformly square-integrable: for some  $\epsilon > 0$ ,*

$$\sup_{n \in \mathbb{N}} \mathbf{E} [ \|\psi(i, \mathcal{N}_n, G, W, r_n)\|^{2+\epsilon} ] < \infty. \quad (10)$$

*Further, assume that for any  $a \in \mathbb{R}^+$ ,  $b \in \mathbb{R}^d$ , with probability one*

$$\psi(i, \mathcal{N}, G, W, r) = \psi(ai + b, a\mathcal{N} + b, G, W, ar). \quad (11)$$

*Then under Assumptions 1-7,  $\frac{1}{n} \sum_{i \in \mathcal{N}_n} \psi(i, i + r_n^{-1}(\mathcal{N}_n - i), G, W, 1) \xrightarrow{L_2} \mu$ , where*

$$\mu = \int \mathbf{E} [ \psi(i, \mathcal{P}_{\kappa f(i)}, G, W, 1) \mid i ] f(i) di.$$

<sup>12</sup>The path distance between nodes  $i$  and  $j$  is the number of links in the shortest path connecting them.

<sup>13</sup>Note that the type shares of their continuum model are different than the limits derived under our theory, since in our limit model, the set of nodes is countable.

The theorem states that, after rescaling the set of node positions *relative to*  $i$  (the first input of each node statistic) by a dilation factor of  $r_n^{-1}$ , network moments on  $\mathcal{N}_n$  converge to their analogs applied to a Poisson process. The relative dilation has the effect of reducing the density of points in any neighborhood of  $i$  by a factor of approximately  $\frac{n}{\kappa}$ , so that the distribution of points is asymptotically  $\mathcal{P}_{\kappa f(i)}$ . This is useful because Poisson processes possess a well-known spatial independence property, as we discuss in the sketch of the proof below.

Condition (10) is a regularity condition obviously satisfied by bounded node statistics. The moments discussed in §6 obey this assumption. Condition (11) states that node statistics are *stationary* in the sense of being invariant to dilation and translation of node positions. In the appendix §A.1, we show that it is essentially sufficient to assume that the primitives  $\mathcal{S}$  and  $\lambda$  depend on node positions  $\mathcal{N}$  only through *effective dissimilarities*  $\{r^{-1}\|i - j\|; i, j \in \mathcal{N}\}$ . This condition is entirely innocuous when positions are merely labels, as with Erdős-Rényi opportunity graphs. If positions correspond to homophilous attributes, as with random geometric graphs or the random connections model, the condition states that the model primitives do not depend on the absolute values of node positions but rather on their relative values. This is natural when position corresponds to geographic location, but if position includes characteristics such as income, this condition is clearly stronger.

**Example 15.** To understand  $\mu$ , consider the simple example in which the opportunity graph is an RGG, and  $\psi_i(\mathcal{N}_n, r_n) = \sum_j \mathbf{1}\{\|i - j\| \leq r_n\}$ , which is node  $i$ 's degree with respect to the opportunity graph. As shown in Example 8,  $\frac{1}{n} \sum_{i \in \mathcal{N}_n} \psi_i(r_n)$  converges to the expected degree, which equals

$$\mathbf{E} \left[ \frac{1}{n} \sum_{i,j} \mathbf{1}\{\|i - j\| \leq r_n\} \right] = \int nr_n^d \text{Vol}(B(0, 1)) f(i)^2 di + o(1). \quad (12)$$

We can check that the right-hand side equals  $\mu$ . Notice

$$\psi(i, \mathcal{N}_n, G, W, r_n) = |\{j \in \mathcal{P}_{\kappa f(i)} : \|i - j\| \leq 1\}|.$$

Therefore,

$$\begin{aligned} \mu &= \int \mathbf{E} [|\{j \in \mathcal{P}_{\kappa f(i)} : \|i - j\| \leq 1\}| | i] f(i) di \\ &= \int \int_{j: \|i-j\| \leq 1} \kappa f(i) dj f(i) di \\ &= \kappa \text{Vol}(B(0, 1)) \int f(i)^2 di, \end{aligned}$$

where the second line follows from a basic property of Poisson processes (Kingman, 1992, Campbell's theorem). The last line is precisely (12) up to an  $o(1)$  term if  $nr_n^d \rightarrow \kappa$ , as required by Assumption 6.

**Proof Sketch.** Let  $\psi_i(\mathcal{N}, r) \equiv \psi(i, \mathcal{N}, G, W, r)$ . Then

$$\begin{aligned} \frac{1}{n} \sum_{i \in \mathcal{N}_n} (\psi_i(\mathcal{N}_n, r_n) - \mu) &= \underbrace{\frac{1}{n} \sum_{i \in \mathcal{N}_n} (\psi_i(\mathcal{N}_n, r_n) - \mathbf{E}[\psi_i(\mathcal{N}_n, r_n) | W, \mathcal{N}_n])}_{[I]} \\ &\quad + \underbrace{\frac{1}{n} \sum_{i \in \mathcal{N}_n} (\mathbf{E}[\psi_i(\mathcal{N}_n, r_n) | W, \mathcal{N}_n] - \mu)}_{[II]}. \end{aligned} \quad (13)$$

Convergence of [I] follows from ideas (A) and (B). Specifically, Assumptions 2 and 5 imply that for any  $C, C' \in \mathcal{C}^+(D, \mathcal{N}_n, r_n)$ ,  $C \neq C'$ , it is the case that

$$\{G_{ij} : i, j \in C\} \perp \{G_{ij} : i, j \in C'\} | W, \Pi(r_n), \mathcal{N}_n \quad (14)$$

For instance, for  $i, j \in C$ ,  $G_{ij}$  only depends on the attributes and positions of nodes in  $C$  by Assumption 2, and nodes in  $C, C'$  independently coordinate on equilibrium subnetworks by Assumption 5.

While links are conditionally independent, the object is to prove a weak law for averages of node statistics, which aggregate across links. In general, node statistics may still be conditionally dependent because the locality restriction allows for dependence on indirect links. For example, in Figure 3, node 4's node statistic may depend on node 1's equilibrium subnetwork through  $G_{23}$ . Sparsity of the opportunity graph will be crucial for limiting this form of dependence, since a node's degree in the opportunity graph is almost-surely finite in the limit under sparsity. Together with subcriticality of  $D$ , it will follow that, conditional on  $D$  and  $\mathcal{N}_n$ , node statistics are dependent with at most a finite number of other node statistics in the limit. Hence, node statistics are only weakly conditionally dependent.

Formally, to prove convergence of [I], it suffices to show concentration of the conditional variance of [I] at zero. Define

$$J_i = C_i^+(D) \cup \{j \in \mathcal{N} : \Pi_{kl}(r_n) = 1 \text{ for some } k \in C_i^+(D), l \in C_j^+(D)\}.$$

This is the set of nodes whose node statistics are potentially conditionally dependent with  $i$ 's node statistic. For instance, in Figure 5,  $C_1^+ = \{1, 2, 3, 4\}$ ,  $C_9^+ = \{6, 7, 8, 9, 10\}$ , and  $\Pi_{46}(r_n) = 1$ , so  $9 \in J_1$ . Notice that 1 and 9's node statistics are conditionally dependent because they both may depend on  $Z_6$ , for instance. On the other hand,  $C_{10}^+ = \{9, 10\}$ , so  $10 \notin J_1$ , and the model assumptions imply that the node statistics of 1 and 10 are conditionally independent.

In general, if  $j \notin J_i$ , then (14) implies that

$$\psi_i(\mathcal{N}_n, r_n) \perp \psi_j(\mathcal{N}_n, r_n) | W, \mathcal{N}_n.$$

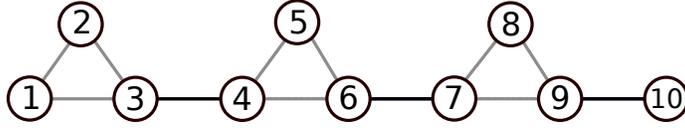


Figure 5: The black links are exogenously formed,  $C_1^+ = C_2^+ = C_3^+ = \{1, 2, 3, 4\}$ ,  $C_4^+ = \{3, 4, 5, 6, 7\}$ ,  $C_7^+ = \{6, 7, 8, 9, 10\}$ ,  $C_{10}^+ = \{9, 10\}$ .

We can then bound the conditional variance of  $[I]$  as follows using (10):

$$\begin{aligned} \text{Var}([I] | D(r_n), \mathcal{N}_n) &= \frac{1}{n^2} \sum_{i \in \mathcal{N}_n} \sum_{j \in J_i} \text{Cov}(\psi_i(\mathcal{N}_n, r_n), \psi_j(\mathcal{N}_n, r_n) | W, \mathcal{N}_n) \\ &\leq \underbrace{\frac{c}{n} \frac{1}{n} \sum_{i=1}^n |J_i|}_{[III]}. \end{aligned}$$

It now remains to show that (A)  $[II]$  and  $[III]$  converge to their expectations, and (B) the limit of  $[III]$  is finite. Part (B) follows because components in the limit model are almost surely finite by Proposition 1, and each node has an opportunity link with an almost surely finite number of other nodes by sparsity (Assumption 6).

For part (A), notice that both of these quantities can be written as averages of functionals of  $W, \mathcal{N}_n$ . Convergence of these functionals will follow from Theorem 3 in §A.2, which extends Theorem 2.1 of Penrose and Yukich (2003) (PY). PY's result yields a weak law for functionals of class of geometric graphs. The intuition behind their result is that the binomial point process  $i + r_n^{-1}(\mathcal{N}_n - i)$  is locally Poisson in a neighborhood of  $i$ . This is useful because unlike binomial point processes, Poisson processes possess a spatial independence property. Notice that the definition of the binomial point process requires translating and dilating node positions, which is why (11) is needed.

To be more specific, a coupling argument shows that the rescaled expectation  $\mathbf{E}[\psi_i(i + r_n^{-1}(\mathcal{N}_n - i), 1) | W, i + r_n^{-1}(\mathcal{N}_n - i)]$  can be locally approximated by its analog applied to a Poisson point process. If the resulting functionals satisfy a stability property, this yields distributional convergence, and a weak law then follows from uniform integrability. The idea behind the stability property is that for the local approximation to work, there must exist a finite radius  $R$  such that changes to the network beyond the ball of radius  $R$  centered at  $i$  do not alter this conditional expectation. Because  $|J_i|$  is finite by Proposition 1 and Assumption 6, we can simply choose  $R$  large enough to contain  $|J_i|$ , and stability follows.

**Remark 4.** Theorem 3 establishes a weak law for functionals of a large class of random graph models, which may be of independent interest. The theorem in PY does

not cover the RCM. Note that the RCM encompasses RGGs, Erdős-Rényi graphs, and SBMs.

## 5.2 Network Properties

Under Assumptions 3 and 6, the network  $G$  is sparse. We next show that networks generated by the model may *percolate*, or contain a giant component (Barabási, 2015; Jackson, 2010). Formally,  $G$  percolates if the size of its largest component is of asymptotic order  $n$  with probability approaching one. The results in this section are informed by the following observations.

- (A) For  $G$  to percolate it is necessary that  $\Pi$  percolates, since  $G$  is a subgraph of  $\Pi$  under Assumption 3.
- (B) Heuristically, if enough pairwise-stable subnetworks formed on each component of  $D$  are sufficiently connected, and enough of these subnetworks are linked together by exogenously formed links, then  $G$  has a giant component.
- (C) Let  $\Pi^-$  be the subnetwork of  $\Pi$  consisting of solely of exogenously formed links. That is, link  $i$  and  $j$  if and only if  $\inf_s V(s, W_{ij}; \theta_0) > 0$ . If  $\Pi^-$  percolates, then it is a subgraph of  $G$ ,  $G$  must also percolate.

In what follows, we will establish conditions for  $\Pi(r_n)$  and  $\Pi^-(r_n)$  to percolate for each of the four examples of opportunity graphs discussed in §4.1. In light of (A) and (C), these are, respectively, necessary and sufficient conditions for  $G$  to percolate. These sufficient conditions are likely stronger than conditions that establish (B). However, it is difficult to derive general primitive restrictions on  $V$  that ensure connectivity of equilibrium subnetworks. In contrast, approach (C) is mathematically feasible due to the availability of relevant results in random-graph theory.

In what follows, it will be useful to define

$$\gamma = \operatorname{ess\,sup}_{i, Z_i} \kappa f(i) \int \mathbf{P}(E_{ij}^c(1) \mid i, Z_i, j, \Pi_{ij}(1) = 1) dj.$$

Note that  $\gamma$  is an upper bound on the probability that a link is not exogenously realized, given that they have an opportunity to link. We will see that percolation typically requires  $\kappa$  to lie in a bounded interval and  $\gamma < 1$ , meaning that a nontrivial share of links must be exogenously realized.

**Example 16** (Random Geometric Graph). Let  $\Pi(r_n)$  be the random geometric graph that links  $i, j \in \mathcal{N}_n$  if and only if  $\|i - j\| \leq r_n$ . Suppose that  $f$  is bounded away from zero on its support,  $\underline{f} = \operatorname{ess\,inf}_x f(x)$ , and

$$\kappa \underline{f} > T, \tag{15}$$

where  $T$  is the ‘‘continuum percolation threshold’’ (Penrose, 2003). This is the threshold above which RGGs defined on a Poisson point process with constant intensity percolate and below which they do not.<sup>14</sup> Under this condition,  $\Pi(r_n)$  has a giant component.

$\Pi(r_n)$  percolation. Let  $\vartheta(f) = \bar{f} \text{Vol}(B(0, 1))$ . A sufficient condition for Assumption 7 is<sup>15</sup>

$$\kappa\gamma\vartheta(f) < 1. \tag{16}$$

Equations (15) and (16) imply that

$$T/\underline{f} < \kappa < (\vartheta(f)\gamma)^{-1},$$

which is satisfied if  $\gamma < \underline{f}(T\vartheta(f))^{-1}$ . The right-hand side is less than  $\underline{f}/\bar{f}(T\bar{\vartheta})^{-1}$ , where  $\bar{\vartheta}$  is the volume of the unit ball. Since  $T \geq 1$  for percolation, and  $\bar{\vartheta} \geq 1$ , it follows that  $\gamma < 1$  is necessary.

$\Pi^-(r_n)$  percolation. We claim that  $\Pi^-(r_n)$  percolates if

$$\text{ess inf}_{i,z} \gamma^*(i, z)\kappa\underline{f} > T, \tag{17}$$

where

$$\gamma^*(i, z) = \inf_n \text{ess inf}_{j,z'} \mathbf{P} \left( \inf_s V(s, W_{ij}; \theta_0) > 0 \mid i, j, Z_i = z, Z_j = z', \Pi_{ij}(r_n) = 1 \right),$$

supp( $f$ ) meaning the support of  $f(\cdot)$ . To see this, define the RGG  $\mathcal{R}(r_n)$ , where node positions are drawn from the density  $\gamma^*(i, z)f(i)d\mu(i, z)$  with  $\mu$  the distribution of  $(i, Z_i)$ . This is a subgraph of the network formed by deleting nodes (and their links) from  $\Pi(r_n)$  with probability  $1 - \gamma^*(i, Z_i)$  for each  $i$ . Hence,  $\Pi^-(r_n)$  can be coupled to  $\mathcal{R}(r_n)$  such that the former is a subgraph of the latter. Since the expected degree of  $\Pi^-(r_n)$  is  $\gamma^*\kappa\underline{f}$ , the graph percolates by Penrose (2003) Theorem 10.9.

To understand the usefulness of (17), consider the following specification for the joint surplus:

$$\theta_1\rho(i, j) + S'_{ij}\theta_2 + \zeta_{ij}, \tag{18}$$

where  $\rho$  is defined in (2) and  $\zeta_{ij}$  is standard normal for any  $i, j \in \mathbb{R}^d$ . If  $S'_{ij}\theta \geq 0$ , which is the case when externalities are positive, then  $\gamma^* = 0.5$ , and  $\gamma = \mathbf{P}(-S'\theta \leq \zeta_{ij} \leq 0)$ , where  $\bar{S}$  is the largest value of  $S_{ij}$  on its support. If  $d = 2$  and positions are uniformly distributed on  $[0, 1]^2$ , then (16) and (17) imply

$$2T < \kappa < (\gamma\pi)^{-1},$$

<sup>14</sup>A fundamental result in continuum percolation states that  $T \in (0, \infty)$  if  $d \geq 2$ . There is no analytical expression for  $T$ ; for  $d = 2$  and  $f$  uniform, simulations indicate that this threshold is approximately 1.44 (Penrose, 2003, p. 189).

<sup>15</sup>We conjecture that (16) can be relaxed to  $\kappa\gamma\vartheta(f) < T$ . Simulation results appear to confirm this conjecture for the case  $d = 2$ . Formally establishing this tight upper bound on  $\kappa\gamma\vartheta(f)$  requires new results on the percolation threshold for edge percolation on random geometric graphs, which is beyond the scope of this paper.

which requires  $\gamma \leq (2T\pi)^{-1}$ , or more simply  $\gamma \leq (2\pi)^{-1}$  under the conjecture of footnote 15. Hence,  $\gamma < 1$  is necessary.

**Example 17** (Stochastic Block Model). Note that the ER model is special case of the SBM with a single group. For simplicity, suppose the model is given by (6) and  $\Pi_{ij} \perp \tilde{V}(s, W_{ij}; \theta_0)$ .

$\Pi(r_n)$  percolation. By Theorem 3.1 of Bollobás et al. (2007), for the model in Example 7,  $\Pi(r_n)$  percolates if

$$\rho^* \equiv \mathbf{E}[\rho(\alpha_i, \alpha_j)] > 1, \quad (19)$$

i.e. the expected degree exceeds one. A sufficient condition for Assumption 7 is

$$\gamma \rho^* \bar{f} < 1 \quad (20)$$

(recall  $\kappa = 1$  in this model). Then (19) and (20) are consistent if

$$1 < \rho^* < (\gamma \bar{f})^{-1},$$

so a necessary condition is  $\gamma \leq \bar{f}^{-1}$ .

$\Pi^-(r_n)$  percolation. By Theorem 3.1 of Bollobás et al. (2007) the graph  $\Pi^-(r_n)$  percolates if

$$\mathbf{P} \left( \inf_s V(s, W_{ij}; \theta_0) > 0 \right) \rho^* > 1, \quad (21)$$

meaning that the expected degree of  $\Pi^-(r_n)$  exceeds one.

Clearly (21) and (20) are consistent if

$$1 < \rho^* < \left( \mathbf{P} \left( \inf_s V(s, W_{ij}; \theta_0) > 0 \right)^{-1} \gamma \bar{f} \right)^{-1}.$$

Then a necessary condition is  $\gamma \leq \mathbf{P}(\inf_s V(s, W_{ij}; \theta_0) > 0) \bar{f}^{-1}$ .

**Example 18** (Random Connection Model). For  $\Pi(r_n)$  given by the model in Example 9, to the best of our knowledge, sufficient conditions for percolation do not yet exist. However, graphs generated by the finite model are typically shown to percolate by first showing that outputs of the limit model in which  $\mathcal{N} = \mathcal{P}_{\kappa f(i)}$  percolate and then deriving a suitable coupling to the finite model.<sup>16</sup> Results do exist for percolation for the limit model where  $\mathcal{N} = \mathcal{P}_\tau \cup \{0\}$ ,  $\tau \in (0, \infty)$ . We will restrict attention to showing that the graphs  $\Pi(1)$  and  $\Pi^-(1)$  percolate under the limit model under Example 1 for  $\rho(i, j) = ||i - j||$  and  $\alpha_i = 0$  for all  $i$ .

$\Pi(1)$  percolation. In the general random connection model of Example 3, we have  $\theta_2 = -r^{-1}$ , and  $r = 1$  in the limit model. Hence, we have

$$\Pi_{ij} = \mathbf{1} \{ \theta_1 + \theta_2 ||i - j|| + \theta_3 + \theta_4 \bar{L} + \zeta_{ij} > 0 \}.$$

<sup>16</sup>This is the strategy used by Penrose (2003) to prove that (15) is sufficient for RGGs to percolate.

Let  $F(\|i - j\|) = \mathbf{P}(\theta_1 + \|i - j\| + \theta_3 + \theta_4 \bar{L} + \zeta_{ij} \geq 0 \mid \|i - j\|)$ . Similar to (9), we require sparsity:  $0 < \tau \int_{x \geq 0} F(x) dx < \infty$ , meaning that the expected number of connections to the origin is finite (also see Example 11). Then Theorem 6.1 of Meester and Roy (1996) establishes the existence of a threshold  $\tau^*(F) > 0$  such that  $\Pi(1)$  percolates<sup>17</sup> if

$$\tau > \tau^*(F). \tag{22}$$

An analogous sufficient condition for  $\Pi(r_n)$  would be  $\kappa f(i) > \tau^*(F)$  for all  $i$ , although this is only a conjecture. Next, define

$$\gamma(x) = \mathbf{P}(\theta_1 + \theta_2 x + \zeta_{ij} \leq 0 \mid \theta_1 + \theta_2 x + \theta_3 + \theta_4 \bar{L} + \zeta_{ij} > 0).$$

Then Assumption 7 states that  $\tau \|\int \gamma(\|i - j\|) F(\|i - j\|) dj\|_2 < 1$  (see Example 11 for further discussion). The left-hand side equals  $\int_{x \geq 0} \gamma(x) F(x) dx$ , since  $j$  ranges over  $\mathbb{R}^d$ , so this is compatible with (22) if

$$\tau^*(F) < \tau < \left( \int_{x \geq 0} \gamma(x) F(x) dx \right)^{-1}.$$

The analogous condition for  $\Pi(r_n)$  would be that  $\kappa f(i)$  stays within those bounds for all  $i$ .

$\Pi^-(1)$  percolation. Define

$$\begin{aligned} \Pi^-(1) &= \mathbf{1} \{ \theta_1 + \theta_2 \|i - j\| + \zeta_{ij} > 0 \}, \\ \tilde{F}(x) &= 1 - \mathbf{P}(-\theta_1 - \theta_2 x). \end{aligned}$$

Then there exists  $\tau^*(\tilde{F}) > 0$  such that  $\Pi^-(1)$  percolates if  $\tau > \tau^*(\tilde{F})$ . The analogous condition for  $\Pi(r_n)$  would be  $\kappa f(i) > \tau^*(\tilde{F})$  for all  $i$ . This former condition is compatible with Assumption 7 if

$$\max \{ \tau^*(F), \tau^*(\tilde{F}) \} < \tau < \left( \int_{x \geq 0} \gamma(x) F(x) dx \right)^{-1}.$$

### 5.3 Simulating Counterfactuals

Under the model assumptions, pairwise-stable networks can be simulated quickly. To the best of our knowledge, this is the first model of network formation for which it is computationally feasible to generate counterfactual networks. Given  $\lambda$ ,  $\mathcal{N}$ ,  $W$ , and  $\theta_0$  we propose the following algorithm:

1. Generate  $\Pi(r_n)$ .
2. Construct  $D$  by removing exogenously realized links from  $\Pi(r_n)$ .

---

<sup>17</sup>The definition of percolation in continuum models is that the component of  $\Pi(1)$  containing the origin has a strictly positive chance of containing infinitely many nodes.

## A WEAK LAW FOR NETWORK MOMENTS

3. For each component of  $D$ , generate a pairwise-stable subnetwork, taking as given the exogenously realized links of  $G$ .

In our Monte Carlo, we find that for a network of 1000 nodes, it takes seconds to find a pairwise-stable subnetwork using myopic best-response dynamics in step 3. We next show that the algorithm has complexity  $O_p(n^2)$ , primarily due to the fact that, under certain conditions, component sizes of  $D$  are only order  $\log n$ . However, even without these conditions, this algorithm is faster than the usual method of running myopic best-response dynamics on the entirety of  $\mathcal{N}_n$ .

For a given set of node positions, the first step has complexity  $O(n^2)$  in general.<sup>18</sup> The second step has complexity  $O_p(n)$ , since it requires iterating through every link in  $\Pi(r_n)$ , and the expected number of links is finite by Assumption 6. We next show that the third step has complexity  $O_p(n^2)$ . The following result is key.

**Proposition 2.** *Suppose that*

$$\limsup_{n \rightarrow \infty} \kappa p_n^* < 1, \text{ where} \tag{23}$$

$$p_n^* = \text{ess sup}_{i,j,Z_i,Z_j} \mathbf{P} \left( E_{ij}^c(r_n) \cap \Pi_{ij}(r_n) = 1 \mid i, j, Z_i, Z_j \right).$$

*Then for any  $i \in \mathcal{N}_n$ , it is the case that  $|C(i, \mathcal{N}_n, D, r_n)| = O(\log n)$  with probability tending to one.*

This is the analog of Proposition 1 for the finite model. Equation (23) strengthens Assumption 7 by taking supremums over the conditional probability. It may be possible to relax this condition with a more sophisticated argument.

The third step of the algorithm can be broken down into three parts. (A) It iterates through each component of  $D$ , say,  $C_1, \dots, C_m$ . (B) For a given such component  $C_k$ , the algorithm extracts the subnetwork of  $\Pi$  on  $C_k$ , denoted  $D(C_k)$ . (C) Since links form only between nodes who are linked in  $\Pi$ , and exogenously realized links are always stable, the algorithm need only iterate through every possible subnetwork of  $\Pi(C_k)$  and check its pairwise stability given the exogenously realized links of  $G$ .

Proposition 2 states that any component of  $D$  has size  $O(\log n)$  with high probability. We then claim that there are  $O_p(n)$  possible subnetworks in step (C) when  $|C_k| = O(\log n)$ . To see this, note that by (C), the number of possible stable networks

---

<sup>18</sup>Random geometric graphs can be generated in  $O(n \log n)$  time using k-d trees.

on a set of nodes  $C$  is at most  $\exp \left\{ \sum_{i,j \in C} \Pi_{ij}(r_n) \right\}$ . Further,

$$\begin{aligned} \mathbf{P} \left( \exp \left\{ \sum_{i,j \in C} \Pi(r_n)(i,j) \right\} > n\epsilon \right) \\ = \mathbf{P} \left( \sum_{i,j \in C} \Pi(r_n)(i,j) > \log n + \log \epsilon \right) \\ \leq \frac{n\mathbf{E}[\Pi(r_n)(i,j)]|C|^2/n}{\log n + \log \epsilon}. \end{aligned}$$

By Assumption 6,  $n\mathbf{E}[\Pi(r_n)(i,j)] = O(1)$ . Thus, if  $|C| = O(\log n)$ , the right-hand side tends to zero, so the computational complexity of finding a pairwise-stable sub-network is at most of order  $n$ . Since this search need only be repeated at most  $n$  times in step (A) (in the worst case, there is one component for each node), the complexity of the algorithm's third step is order  $n^2$  on average.

## 6 Identification and Estimation

### 6.1 Moment Inequalities

In this section, we study the identification of  $\theta_0$  by constructing a set of moment inequalities that hold if and only if the data is rationalized by a pairwise-stable equilibrium (i.e. Assumption 4(b) holds). We will impose the following assumptions.

**Assumption 8.** *The range of  $S$ , denoted by  $\Psi$ , is finite.*

This assumption is convenient to impose for computational reasons, and analogous assumptions are used in de Paula et al. (2015) and Sheng (2014). Example 1 satisfies this condition, and it is often simple to modify a surplus function to ensure finiteness. We emphasize that this assumption is not necessary for deriving a useful characterization of the identified set in general. We employ it here to make direct use of a theorem due to Beresteanu et al. (2011) to characterize the empirical content of the model in terms of a finite set of conditional moment inequalities. However, continuously distributed endogenous statistics can easily be accommodated by following our approach below but applying Theorem 1 of Galichon and Henry (2011) to derive a characterization in terms of an infinite set of conditional moment inequalities.

**Assumption 9** (Analyst's Information).

- (a) *The public signal  $\nu$  and selection mechanism  $\lambda$  defined in Assumption 4 and a subvector of  $W_{ij}$  are unobserved. We denote the unobserved subvector by  $\varepsilon_{ij}$  and  $X_{ij} = (W_{ij} \setminus \varepsilon_{ij})$ .*

## A WEAK LAW FOR NETWORK MOMENTS

(b) *The distribution of  $\varepsilon_{ij} | X_{ij}$  is known up to a finite-dimensional parameter. Without loss of generality, this parameter is a subvector of  $\theta_0$ .*

Part (b) is standard. The requirement that  $\nu$  and  $\lambda$  are unobserved in part (a) implies that the equilibrium selection mechanism is unknown, which is the usual case of interest.

Before presenting the theorem, we need several definitions. Define the *dyadic outcome*  $Y_{ij} = (G_{ij}, S_{ij})$ , where  $S$  satisfies Assumption 2. Notice that Assumption 8 ensures that  $\mathcal{Y} = \{0, 1\} \times \Psi$ , the range of  $Y_{ij}$ , is finite. We say  $Y_{ij} = (\ell, s)$  is *stable with respect to  $W_{ij}$ ,  $\theta$ , and  $r$*  if

$$\begin{aligned} V(s, W_{ij}; \theta) &> 0 && \text{if } \ell = 1, \\ V(s, W_{ij}; \theta) &\leq 0 && \text{if } \ell = 0. \end{aligned}$$

(Recall that  $V$  depends on  $r$  under Assumption 3.) The *stable set*  $\mathcal{S}_\theta(W_{ij}, r)$  is the set of dyadic outcomes that are stable under  $W_{ij}$ ,  $\theta$ , and  $r$ . We can then define the random set

$$Q_\theta(W_{ij}, r) = \{(\mathbf{1}\{Y = y\}\Pi_{ij}(r), y \in \mathcal{Y}); Y \in \mathcal{S}_\theta(W_{ij}, r)\}.$$

Note that we multiply by  $\Pi_{ij}(r)$  because pairs of nodes that are unlinked in  $\Pi$  provide no information on  $\theta_0$ , since their equilibrium links never form by Assumption 3. Lastly, define the vector  $\mathbf{Y}_{ij}(r) = (\mathbf{1}\{Y_{ij} = y\}\Pi_{ij}(r), y \in \mathcal{Y})$ .

**Theorem 2.** *Let  $\mathcal{U} = \{0, 1\}^{|\mathcal{Y}|}$ . Under Assumptions 2, 3, 4(a), 8, and 9, for any fixed  $n$ , the observed network is rationalized by a pairwise-stable equilibrium in the sense of Assumption 4(b) if and only if with probability one,*

$$u' \mathbf{E}[\mathbf{Y}_{ij}(r_n) | X_{ij}] \leq \mathbf{E} \left[ \sup_{q \in Q_{\theta_0}(W_{ij}, r_n)} u' q \mid X_{ij} \right] \quad \forall u \in \mathcal{U}. \quad (24)$$

The theorem can be easily generalized to models that do not require opportunity graphs by dropping all appearances of  $\Pi$  in the definitions of the objects above. The principles behind the theorem also readily extend to other large games.

Theorem 2 demonstrates that to characterize the empirical content of the model under unrestricted selection, it is sufficient to consider statistics that take the form of dyadic-outcome moments. The idea is as follows. In the traditional cross-sectional setting in which the econometrician observes a large number of independent networks, we typically conceptualize the model as a mapping from the primitives  $(X, \varepsilon)$  to the set of networks  $G$  that are pairwise stable. This is sensible because cross-sectional data reveals the joint distribution of links. However, this distribution is not revealed if only a single network is observed. Our insight is that in a single-network setting, the econometrician instead observes a large number of dependent dyadic outcomes.

We therefore conceptualize the model as a mapping from the attributes of a dyad  $(X_{ij}, \varepsilon_{ij})$  to the set of possible dyadic outcomes  $Y_{ij}$  that are stable.<sup>19</sup> Theorem 2 establishes that these two characterizations of the model are equivalent and then derives moment inequalities under the second model characterization using Theorem 1 of Beresteanu et al. (2011).

**Example 19.** Consider the specification in Example 1 with  $\theta_4 = 0$ ,  $\rho$  defined in (2), and  $\alpha_i = 0$  for all  $i$ . In this specification,  $\mathcal{Y} = \{0, 1\}^2$ . Notice that for any  $s \in \{0, 1\}$ , exactly one of  $(1, s)$  or  $(0, s)$  is in  $\mathcal{S}_\theta(W_{ij}, r)$ ; we refer to this as observation  $(\star)$ . Consequently the latter set has cardinality two. Let  $u = (u_1, u_2, u_3, u_4) \in \mathcal{U}$ , where we associate  $u_1$  with  $Y_{ij} = (1, 1)$ ,  $u_2$  with  $(0, 1)$ ,  $u_3$  with  $(1, 0)$ , and  $u_4$  with  $(0, 0)$ .

Abusing notation, for  $s, t \in \mathcal{Y}$ , let  $\{y, z\}$  be the event that  $y$  and  $z$  are stable with respect to  $W_{ij}$ ,  $\theta$ , and  $r$  and that the nodes are linked in  $\Pi$ . For example,  $\{(1, 1), (0, 0)\}$  equals

$$\{\theta_1 + \theta_3 + \zeta_{ij} \geq 0\} \cap \{\theta_1 + \zeta_{ij} < 0\} \cap \{\Pi_{ij}(r) = 1\}.$$

From observation  $(\star)$ , it is then easy to see that

$$\begin{aligned} \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, r)} u'q \mid X_{ij} \right] &= \max\{u_1, u_3\} \mathbf{P}((1, 1), (1, 0) \mid X_{ij}) \\ &\quad + \max\{u_1, u_4\} \mathbf{P}((1, 1), (0, 0) \mid X_{ij}) \\ &\quad + \max\{u_2, u_3\} \mathbf{P}((0, 1), (1, 0) \mid X_{ij}) \\ &\quad + \max\{u_2, u_4\} \mathbf{P}((0, 1), (0, 0) \mid X_{ij}). \end{aligned}$$

Note that the conditional probabilities can be easily simulated as the events  $\{y, z\}$  simply define a partition of  $\zeta$ -space, which is one-dimensional. Moreover, if  $\zeta_{ij}$  is independent of observables and, say, normally distributed, then the conditional probabilities can be computed in closed form.

In the next two subsections, we consider different assumptions for the analyst's information set and construct consistent estimators for some of the moments in (24). The availability of consistent estimators then enables us to define the identified set of parameters.

## 6.2 Estimation: Observed Opportunities

We first consider estimation under the following assumption.

---

<sup>19</sup>The focus on the dyad as the unit of observation follows because under (1), we can conceptualize the network-formation process as a game between pairs of nodes taking binary actions. Note that the results in this section can also be easily extended to the case of nontransferrable utility.

**Assumption 10.**  $\Pi$  is known up to  $\theta_0$ .

This is satisfied by random geometric graphs when node positions are observed, since links are then known up to  $r$ . It is also satisfied by some cases of random connection models.

First consider estimation of  $\lim_{n \rightarrow \infty} n \mathbf{E}[u' \mathbf{Y}_{ij}(r_n) h(X_{ij})]$  in (28). The empirical analog is written

$$\frac{1}{n^2} \sum_{i,j \in \mathcal{N}_n} u' n \mathbf{Y}_{ij}(r_n) h(X_{ij}) = \frac{1}{n} \sum_{i \in \mathcal{N}_n} \sum_{\substack{j \in i+r_n^{-1}(\mathcal{N}_n-i), \\ \Pi_{ij}(1)=1}} u' \mathbf{Y}_{ij}(1) h(X_{ij}), \quad (25)$$

where the equality follows from (11).<sup>20</sup> We next apply Theorem 1. Define

$$\psi^*(i, \mathcal{N}_n, G, W, r_n) = \sum_{j \in \mathcal{N}_n} u' \mathbf{Y}_{ij}(r_n) h(X_{ij}).$$

It is straightforward to see that  $\psi^*$  satisfies locality. Using (11), we can write (25) =  $\frac{1}{n} \sum_{i=1}^n \psi^*(i, i + r_n^{-1}(\mathcal{N}_n - i), G, W, 1)$ . Uniform integrability of  $\psi^*$  follows because  $u' \mathbf{Y}_{ij}(r_n) h(X_{ij})$  is uniformly bounded, and by Assumption 6, the expected degree of  $\Pi_{ij}$  is almost surely finite. We can therefore state the following proposition.

**Proposition 3.** *Under the conditions of Theorem 1, for  $u \in \mathcal{U}$ ,*

$$(25) \xrightarrow{p} \kappa \int \int u' \mathbf{E}[\mathbf{Y}_{ij}(1) h(X_{ij}) \mid i, j] f(i)^2 di dj. \quad (26)$$

The limit expression follows from a basic property of Poisson point processes (Kingman, 1992, Campbell's theorem).

Next we turn to estimation of  $\lim_{n \rightarrow \infty} n \mathbf{E}[\sup_{q \in Q_\theta(W_{ij}, r_n)} u' q h(X_{ij})]$ . The empirical analog is

$$\begin{aligned} \frac{1}{n^2} \sum_{i,j \in \mathcal{N}_n} n \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, r_n)} u' q \mid X_{ij} \right] h(X_{ij}) \\ = \frac{1}{n} \sum_{i \in \mathcal{N}_n} \sum_{\substack{j \in i+r_n^{-1}(\mathcal{N}_n-i) \\ \Pi_{ij}(1)=1}} \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, 1)} u' q \mid X_{ij} \right] h(X_{ij}), \end{aligned} \quad (27)$$

where the equality follows under conditions discussed in §A.1. The right-hand side is a feasible estimator because the summands can always be computed via simulation. In some cases a closed form exists, as in Example 19 when  $\zeta_{ij}$  is independent of observables. By the same reasoning for (25), Theorem 1 characterizes the probability limit of this estimator.

<sup>20</sup>Sufficient conditions for (11) are given in §A.1.

**Proposition 4.** *Under the conditions of Theorem 1, for  $u \in \mathcal{U}$ ,*

$$(27) \xrightarrow{p} \kappa \int \int \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, 1)} u'q h(X_{ij}) \mid i, j \right] f(i)^2 di dj.$$

**Identified Set.** In light of Propositions 3 and 4, we can therefore write the identified set defined in (28) explicitly in terms of the limit objects. Let  $\Lambda$  be a distribution over bounded instrument functions  $h$  defined in §3 of Andrews and Shi (2013). Then (24) holds if and only if for any  $i, j \in \mathcal{N}_n$ ,

$$\int \max_{u \in \mathcal{U}} \mathbf{E} \left[ \left( u' \mathbf{Y}_{ij}(r_n) - \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, r_n)} u'q \mid X_{ij} \right] \right) h(X_{ij}) \right] d\Lambda = 0.$$

(cf. Beresteanu et al., 2011). We can initially define the identified set as the set of  $\theta$  and  $r_n$  satisfying

$$\int \max_{u \in \mathcal{U}} \lim_{n \rightarrow \infty} \mathbf{E} \left[ \left( u'n \mathbf{Y}_{ij}(r_n) - n \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, r_n)} u'q \mid X_{ij} \right] \right) h(X_{ij}) \right] d\Lambda = 0. \quad (28)$$

Note that we scale up the expectations by  $n$ , since  $u' \mathbf{Y}_{ij}(r_n) = u' \mathbf{Y}_{ij}(r_n) \Pi_{ij}(r_n)$ , and by Assumption 6, the indicator is  $O_p(n^{-1})$  (also see Example 15 and the discussion of network sparsity in §5.2).

Existence of the limits in (28) follow from the propositions above, which yield the following explicit characterization of the identified set:

$$\Theta_I = \left\{ \theta \in \Theta : \int_h \max_{u \in \mathcal{U}} \kappa \int_i \int_j \mathbf{E} \left[ \left( u' \mathbf{Y}_{ij}(1) - \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, 1)} u'q \mid X_{ij} \right] \right) h(X_{ij}) \mid i, j \right] f(i)^2 di dj d\Lambda = 0 \right\}.$$

If  $X$  has finite support, then the set of instrument functions is finite, and we obtain a reduction to a finite set of unconditional moments. Then a consistent estimator for  $\Theta_I$  can be obtained using Chernozhukov et al. (2007).<sup>21</sup>

**Remark 5** (Estimating  $r_n$  for RGGs). We argue that  $\hat{r} = \max\{\|i - j\| : G_{ij} = 1\}$  is a consistent estimate of  $r_n$  when  $\Pi(r)$  is an RGG. To see this, notice for  $\epsilon > 0$

$$\mathbf{P}(|\hat{r}_n - r_n| > \epsilon r_n) = \mathbf{P}(\hat{r}_n < (1 - \epsilon)r_n)$$

<sup>21</sup>Also see Wan (2013) and Yildiz (2012). Theory for the continuous support case does not exist, to my knowledge. While one could replace the moments in  $\Theta_I$  with sample analogs and perform a grid search, sampling variation could lead to empty sets, even with probability one. A conjecture is that a consistent estimator can be obtained by including parameter values in the set estimator if the empirical analog of the moment in  $\Theta_I$  is less than or equal to  $c \frac{\log n}{n}$  for any  $c > 0$ , the threshold used in Chernozhukov et al. (2007).

because  $\hat{r}_n \leq r_n$  w.p. 1. For  $\epsilon > 1$ , the limit supremum of the RHS is zero. Thus,

$$\lim_{\epsilon \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P}(|\hat{r}_n - r_n| > \epsilon r_n) = 0.$$

If  $\mathcal{N}_n$  is not observed, then an identification at infinity argument can be used to identify  $\frac{1}{n} \sum_{i,j \in \mathcal{N}_n} \mathbf{P}(\|i - j\| \leq r_n)$ , which can be then used to back out  $r_n$  using (8), provided  $f$  is known. Suppose there exists some observed scalar attribute  $X_i$  such that the joint surplus tends to infinity as  $X_i \rightarrow \infty$ . Then  $\lim_{X_i \rightarrow \infty} \mathbf{E} \left[ \frac{1}{n} \sum_{i,j \in \mathcal{N}_n} G_{ij} \mid X_i \right] \approx \frac{1}{n} \sum_{i,j \in \mathcal{N}_n} \mathbf{P}(\|i - j\| \leq r_n)$ . This is because in the limit, node  $i$  is willing to link with any  $j$  such that  $\|i - j\| \leq r_n$ .

### 6.3 Estimation: Unobserved Opportunities

Now we assume that  $\Pi$ , and possibly  $\mathcal{N}_n$ , is unobserved.

**Assumption 11.** *Either  $\Pi_{ij}$  or  $(\Pi_{ij}, i, j)$  is a subvector of  $\varepsilon_{ij}$ , and  $\mu$  is known up to  $\theta_0$ .*

Because  $\Pi_{ij}$  is latent, not all moments characterizing  $\Theta_I$  can be estimated. Clearly (25) and (27) are not observed quantities, since they contain an inner sum that depends on  $\Pi$ . However, (27) is equivalent to

$$\frac{1}{n} \sum_{i,j \in \mathcal{N}_n} \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, 1)} u'q \mid X_{i,j} \right] h(X_{ij}), \quad (29)$$

which is observed. Its expectation can easily be shown to converge to the limit in Proposition 4 using Lemma 3. To establish convergence, it then remains to show that the variance tends to zero. To see this, first notice that the summands equal  $k_n(X_{ij}) \equiv \mathbf{E}[\sup u'q h(X_{ij}) \Pi_{ij} \Pi_{ij}(r_n) \mid X_{ij}]$ , which is uniformly bounded by  $\mathbf{P}(\Pi_{ij}(r_n) \mid X_{ij})$  times a constant. The latter expression is uniformly  $O_p(n^{-1})$  by Assumption 6. Second, the variance equals

$$\frac{1}{n^2} \sum_{i,j \in \mathcal{N}_n} \mathbf{E} [k_n(X_{ij})^2] + \frac{2}{n^2} \sum_{i,j \in \mathcal{N}_n} \sum_{k \neq i} \mathbf{E} [k_n(X_{ij}) k_n(X_{kj})].$$

Since each  $k_n$  is uniformly  $O_p(n^{-1})$ , this is easily seen to be  $O_p(n^{-1})$ .

While (27) can be salvaged, (25) cannot. To see this, order the elements of the vector  $\mathbf{Y}$  such that the first  $|\Psi|$  (recall  $\Psi$  from Assumption 8) elements are associated with  $G_{ij} = 1$  and the last  $|\Psi|$  with  $G_{ij} = 0$ . Let  $\tilde{\mathcal{U}}$  be the set of  $u \in \mathcal{U}$  such that their last  $|\Psi|$  components are equal to zero. Then

$$\frac{1}{n^2} \sum_{i,j \in \mathcal{N}_n} u'n \mathbf{Y}_{ij}(r_n) h(X_{ij}),$$

is only an observed quantity when  $u \in \tilde{\mathcal{U}}$ . This is because by Assumption 3,  $G_{ij} = 1$  only if  $\Pi_{ij} = 1$ , so the presence of  $\Pi_{ij}$  in the definition of  $\mathbf{Y}_{ij}$  is redundant for dyadic outcomes  $(1, s)$ . In other words, only moments of the form

$$\frac{1}{n} \sum_{i,j \in \mathcal{N}_n} \mathbf{1}\{Y_{ij} = (\ell, s)\} \Pi_{ij}(r_n) \quad (30)$$

for  $\ell = 1$  can be estimated, as far as the asymptotic theory in this paper is concerned. However, even though a weak law generally does not apply for  $\ell = 0$ , notice that (30) is bounded above by the empirical average degree of  $\Pi(r_n)$ . By Assumption 6, the limit of its expectation is  $nr_n^d \int \int p_{ij}(1) f(i)^2 di dj$ , which is known up to  $\theta_0$  under Assumption 11. We can therefore replace (30) when  $\ell = 0$  with this upper bound when defining the identified set.

Define

$$M_{ij}(u, r) = \begin{cases} u' \mathbf{Y}_{ij}(r) & \text{if } u \in \tilde{\mathcal{U}} \\ p_{ij}(1) \|u\|_1 & \text{if } u \in \mathcal{U} \setminus \tilde{\mathcal{U}} \end{cases}$$

where  $\|\cdot\|_1$  is the  $l_1$  norm. The identified set is then

$$\Theta_I = \left\{ \theta \in \Theta : \int_h \max_{u \in \mathcal{U}} \kappa \int_i \int_j \mathbf{E} \left[ \left( M_{ij}(u, 1) - \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, 1)} u' q \mid X_{ij} \right] \right) h(X_{ij}) \mid i, j \right] f(i)^2 di dj d\Lambda = 0 \right\}.$$

## 7 Monte Carlo

**DGP.** We conduct a simulation study to illustrate the informativeness of the identified set. We consider the specification

$$\theta_1 + \theta_2 \mathbf{1}\{Z_i \neq Z_j\} + \theta_3 \mathbf{1}\{\exists k : G_{ik} = G_{jk} = 1\} + \rho(i, j) + \zeta_{ij},$$

where  $\rho$  is given by (2), positions are uniformly distributed on  $[0, 1]^2$ ,  $Z_i \stackrel{iid}{\sim} \text{Ber}(0.35)$ ,  $\zeta_{ij} \stackrel{iid}{\sim} N(0, \theta_4)$ , and  $Z \perp \zeta$ . We normalize  $\Theta$  to be the unit ball centered at the origin (with  $\theta_4$  strictly positive) and set  $\theta_0 = (0.2, -0.2, 0.2, 0.8)$ . Then  $\theta_2$  captures homophily in  $Z_i$  and  $\theta_3$  transitivity. The econometrician does not observe  $\zeta$  but observes all other variables.

Due to our choice of  $\rho$ ,  $\Pi$  is a random geometric graph (Example 8). We choose  $\kappa$  to satisfy Assumption 7:

$$\kappa = (\bar{f} \|\gamma_1(i, Z_i)\|_2)^{-1} - 0.0001. \quad (31)$$

We then set  $r = (\kappa n^{-1})^{1/2}$  in accordance with Assumption 6.

## A WEAK LAW FOR NETWORK MOMENTS

We simulate<sup>22</sup> and estimate the model for the cases in which  $r$  is known and  $r$  is estimated according to Remark 5. The results are virtually identical, since  $\hat{r}$  tends to be extremely close to  $r$ . Below we present results only for the  $\hat{r}$  case, which, as noted in Remark 5, can lead to wider estimated bounds relative to the  $r$ -known case.

**Computation.** Since  $X_{ij} = (Z_i, Z_j)$ , we define the set of instrument functions as  $\mathcal{H} = \{\mathbf{1}\{\cdot \neq \cdot\}, \mathbf{1}\{\cdot = \cdot\}\}$ . Since there are a finite number of instruments, we have a finite number of unconditional moment inequalities, and we can write the identified set as

$$\{\theta \in \Theta : m(\theta, u, h) \leq 0 \forall u \in \mathcal{U}, h \in \mathcal{H}\}, \text{ where}$$

$$m(\theta, u, h) = \kappa \int_i \int_j \mathbf{E} \left[ \left( u' \mathbf{Y}_{ij}(1) - \mathbf{E} \left[ \sup_{q \in Q_\theta(W_{ij}, r_n)} u' q \mid X_{ij} \right] \right) h(X_{ij}) \mid i, j \right] f(i)^2 di dj.$$

We estimate the moments, denoted by  $m(\theta, u, h)$ , using the estimators proposed in §6.2 to obtain  $\hat{m}_n(\theta, u, h)$ . The estimated identified set  $\hat{\Theta}_I$  is computed by performing a grid search on  $\Theta$  and including all parameters  $\theta$  that satisfy

$$\sum_{u \in \mathcal{U}} \sum_{h \in \mathcal{H}} \max \{ \hat{m}_n(\theta, u, h), 0 \}^2 \leq 0.0001 \frac{\log n}{n},$$

which follows Chernozhukov et al. (2007). The step size for the grid is 0.2.

**Network Statistics.** Tables 1-3 display summary statistics aggregated over fifty simulated networks with  $n = 5000$ . The average degree is obtained by dividing the number of links by  $n$ . From this quantity, it is clear that the networks are sparse. “Frac. Giant” is the fraction of nodes lying in the giant component. We can see that the networks  $G$  and  $\Pi$  percolate, as conjectured, and virtually all nodes lie within the giant component. As Proposition 7 predicts,  $D$  does not percolate and contains 400 components on average. “Clustering” is the clustering coefficient (Example 13), and we see that  $G$  has a nontrivial amount of clustering.<sup>23</sup>

**Results.** We simulate the model and compute the identified set thirty times. In all but one simulation, the true parameter is in the estimated identified set. Moreover, projections of the estimated set onto individual coordinates of  $\theta$  always contain the corresponding projection of  $\theta_0$ . Table 4 summaries of these projections aggregated across the simulations. The “Mean Endpoints” column displays a set  $[L, U]$  where  $L$  is the smallest value of the parameter in the given dimension, averaged across the ten simulations, and  $U$  is the largest value. The third and fourth columns display, respectively, the narrowest and widest projections across simulations. From the table, we can see that the homophily and variance parameters are tightly estimated. On

<sup>22</sup>The network is simulated using the algorithm detailed in §5.3. We compute a pairwise-stable equilibrium subnetwork on each augmented component of  $D$  using myopic best-response dynamics, starting at the subnetwork of  $\Pi$  on that component.

<sup>23</sup>By comparison, a network generated by an Erdős-Rényi null model has clustering coefficient equal to the fraction of linked pairs, which is essentially zero due to sparsity.

Table 1: Summary statistics for  $G$ .

	Mean	SD	Min	Max
# Links	17717.40	142.72	17386.00	17972.00
Clustering	0.17	0.00	0.17	0.17
Frac. Giant	1.00	0.00	0.99	1.00
# Components	14.62	3.26	7.00	22.00
# Het. Links	7475.12	108.27	7252.00	7657.00

$n = 5000$ , 50 simulations.

Table 2: Summary statistics for  $D$ .

	Mean	SD	Min	Max
# Links	2429.52	42.68	2333.00	2501.00
Clustering	0.02	0.00	0.01	0.03
Frac. Giant	0.01	0.00	0.01	0.01
# Components	2647.44	36.81	2571.00	2728.00

$n = 5000$ , 50 simulations.

the other hand, the set for  $\theta_1$  typically spans a wide range. Since the estimates of the transitivity parameter  $\theta_3$  are always equal to  $0.4 - \theta_1$ , the sign of this parameter is unidentified for most of the simulations, although in some simulations, the largest value of  $\theta_1$  is 0.2, in which case the sign of  $\theta_3$  is correctly estimated.

We obtain similar results across thirty simulations when  $n = 1000$  (for both  $r$  estimated and known). The only substantive difference is that the narrowest set for the variance is  $[0.60, 0.60]$  because in twelve simulations, the estimated identified set does not contain the true parameter. However, these sets do contain  $(0.2, -0.2, 0.2, 0.6)$ , so only the estimate of  $\theta_4$  is off by 0.2. This is likely due to sampling variation, as this phenomenon seems to disappear when  $n = 5000$ .

Table 3: Summary statistics for II.

	Mean	SD	Min	Max
# Links	28230.80	164.55	27920.00	28561.00
Clustering	0.32	0.00	0.32	0.33
Frac. Giant	1.00	0.00	1.00	1.00
# Components	1.54	0.64	1.00	4.00

$n = 5000$ , 50 simulations.

Table 4: Estimate of identified set.

	Mean Endpoints	Narrowest	Widest
Intercept	[-0.60, 0.72]	[-0.60, 0.20]	[-0.60, 1.00]
Homophily	[-0.2]	[-0.2]	[-0.2]
Transitivity	[0.4 - $\theta_1$ ]	[0.4 - $\theta_1$ ]	[0.4 - $\theta_1$ ]
Variance	[0.62, 0.93]	[0.80, 0.80]	[0.60, 1.00]

Grid step size = 0.2,  $n = 5000$ , 30 simulations.

## 8 Conclusion

This paper develops asymptotic theory for network-formation models when the econometrician observes a single network. We derive conditions under which a weak law holds for a class of network moments, which we apply to construct consistent estimators of the identified set characterized by a new set of computationally tractable moment inequalities. We also study the asymptotic properties of networks generated by our model, establishing conditions for sparsity and percolation. Lastly, we propose a fast algorithm for simulating counterfactual networks.

The theory developed in this paper can be easily applied to games on large networks. Similar to the intuition discussed in §3, in this setting, conditional independence between two agents' actions holds if some agent on any path connecting them is hit with a sufficiently large random utility shock. Our arguments for deriving the identified set are also easily applied to other large games.

The asymptotic theory in this paper focuses on establishing a weak law. A central limit theorem may be attainable using results in geometric probability that build on the ideas of [Penrose and Yukich \(2003\)](#) utilized in this paper. We are currently studying this issue in a separate project.

# A Appendix

## A.1 Stationarity Conditions

We provide primitive conditions under which (11) holds. The primary issue is that if  $\psi$  depends nontrivially on  $G$ , equation (11) implies a similar invariance property must be satisfied by the network.

**Definition 2.** Define  $\mathcal{H}$  to be the set of random, vector-valued functions  $H(i, j, \mathcal{S}, r)$  defined for all  $\mathcal{S} \subseteq \mathbb{R}^d$ ,  $i, j \in \mathcal{S}$ , and  $r \in \mathbb{R}^+$ . We say that  $H \in \mathcal{H}$  is *stationary* if  $H(i, j, \mathcal{S}, r) = H(ai + b, aj + b, a\mathcal{S} + b, ar)$  for any  $a \in \mathbb{R}^+$ ,  $b \in \mathbb{R}^d$ .

We can view  $W_{ij}$  as a mapping from  $i, j \in \mathcal{N}$  to a random vector. In follows, we let  $W$  be an element of  $\mathcal{H}$  such that  $W(i, j, \mathcal{N}, r) = (Z_i, Z_j \zeta_{ij})$  for all  $i, j \in \mathcal{N}$ , with  $Z_i$  and  $\zeta_{ij}$  allowed to be  $r$ -dependent.

Note that any network  $\Gamma$  depends implicitly on  $\mathcal{N}$  by definition and on  $r$  by Assumption 3. Hence, networks are elements of  $\mathcal{H}$ . We next provide primitive restrictions under which  $G$  is stationary.

**Assumption 12** (Invariance of  $S, \mu$ ). *The following conditions hold for any  $\mathcal{N} \subseteq \mathbb{R}^d$ ,  $i, j \in \mathcal{N}$ ,  $a, r \in \mathbb{R}^+$ , and  $b \in \mathbb{R}^d$ .*

- (a)  $W$  is stationary.
- (b)  $\mu$  satisfies  $\mu(a\|i - j\|, W_{ij}; ar) = \mu(\|i - j\|, W_{ij}; r)$ .
- (c) For any network  $\Gamma$  on  $\mathcal{N}$  and stationary  $W$ ,

$$\begin{aligned} S(i, j, \Gamma(\cdot, \cdot, \cdot, r), W(\cdot, \cdot, \cdot, r), \mathcal{N}) \\ = S\left(ai + b, aj + b, \Gamma\left(\frac{\cdot - b}{a}, \frac{\cdot - b}{a}, \mathcal{N}, \frac{r}{a}\right), W(\cdot, \cdot, \cdot, ar), a\mathcal{N} + b\right). \end{aligned}$$

Part (a) and (b) impose stationarity on  $\mu$  and  $W$ , respectively. Invariance of  $\mu$  clearly holds for the examples in §4.1. Part (c) is simply an anonymity restriction on  $S$ . We show that this invariance assumption implies a similar invariance condition holds for any equilibrium network.

**Lemma 1.** *Fix any  $\mathcal{N} \subseteq \mathbb{R}^d$ ,  $i, j \in \mathcal{N}$ ,  $a, r \in \mathbb{R}^+$ , and  $b \in \mathbb{R}^d$ . Under Assumption 12, for any  $\Gamma(\cdot, \cdot, \cdot, \mathcal{N}, r) \in \mathcal{G}_{\theta, r}(W, \mathcal{N})$ , there exists  $\Gamma'(\cdot, \cdot, a\mathcal{N} + b, ar) \in \mathcal{G}_{\theta, ar}(W, a\mathcal{N} + b)$  such that*

$$\Gamma(\cdot, \cdot, \mathcal{N}, r) = \Gamma'(\cdot, \cdot, a\mathcal{N} + b, ar). \quad (32)$$

*Likewise, for any  $\Gamma'(\cdot, \cdot, a\mathcal{N} + b, ar) \in \mathcal{G}_{\theta, ar}(W, a\mathcal{N} + b)$ , there exists  $\Gamma(\cdot, \cdot, \mathcal{N}, r) \in \mathcal{G}_{\theta, r}(W, \mathcal{N})$  such that (32) holds.*

## A WEAK LAW FOR NETWORK MOMENTS

PROOF. We prove the first part of the lemma. Let  $\Gamma(\cdot, \cdot, \mathcal{N}, r) \in \mathcal{G}_{\theta, r}(W, \mathcal{N})$ . For any  $\tilde{i}, \tilde{j} \in a\mathcal{N} + b$ , construct

$$\Gamma'(\tilde{i}, \tilde{j}, a\mathcal{N} + b, ar) = \Gamma\left(\frac{\tilde{i} - b}{a}, \frac{\tilde{j} - b}{a}, \mathcal{N}, r\right)$$

for any  $a \in \mathbb{R}^+$ ,  $b \in \mathbb{R}$ . The right-hand side is well defined by construction of  $\tilde{i}, \tilde{j}$ . Then by definition,  $\Gamma'(\cdot, \cdot, a\mathcal{N} + b, ar) \in \mathcal{G}_{\theta, ar}(W, a\mathcal{N} + b)$  if for every  $\tilde{i}, \tilde{j} \in a\mathcal{N} + b$ ,

$$\Gamma'(\tilde{i}, \tilde{j}) = 1 \Leftrightarrow V\left(S(\tilde{i}, \tilde{j}, \Gamma'(\cdot, \cdot, \cdot, ar), W(\cdot, \cdot, \cdot, ar), a\mathcal{N} + b), W(\tilde{i}, \tilde{j}, a\mathcal{N} + b, ar); \theta_0\right) > 0. \quad (33)$$

By construction of  $\Gamma'$ , the joint surplus in the previous display equals

$$V\left(S\left(\tilde{i}, \tilde{j}, \Gamma\left(\frac{\cdot - b}{a}, \frac{\cdot - b}{a}, \mathcal{N}, r\right), W(\cdot, \cdot, \cdot, ar), a\mathcal{N} + b\right), W(\tilde{i}, \tilde{j}, a\mathcal{N} + b, ar); \theta_0\right).$$

Let  $i, j \in \mathcal{N}$  such that  $ai + b = \tilde{i}$  and  $aj + b = \tilde{j}$ . By Assumption 12, the previous expression is equivalent to

$$V\left(S(i, j, \Gamma(\cdot, \cdot, \mathcal{N}, r), W(\cdot, \cdot, \cdot, r), \mathcal{N}), W(i, j, \mathcal{N}, r); \theta_0\right),$$

But by (1), the  $\Gamma(i, j) = 1$  if and only if the previous equation exceeds zero. This establishes (33). A similar argument proves the second statement of the lemma.  $\blacksquare$

As a consequence of this lemma, there exists a bijection  $\phi(\cdot, a, b) : \mathcal{G}_{\theta, r}(W, \mathcal{N}) \rightarrow \mathcal{G}_{\theta, ar}(W, a\mathcal{N} + b)$  such that  $\Gamma(\cdot, \cdot, \mathcal{N}, r) = \phi(\Gamma(\cdot, \cdot, \mathcal{N}, r), a, b)$ .

**Assumption 13** (Invariance of  $\lambda$ ). *For any  $\mathcal{N} \subseteq \mathbb{R}^d$ ,  $i, j \in \mathcal{N}$ ,  $a, r \in \mathbb{R}^+$ , and  $b \in \mathbb{R}^d$ , if the conclusion of Lemma 1 holds, then the selection mechanism  $\lambda$  satisfies*

$$\lambda_{\theta_0, ar}(W, a\mathcal{N} + b, \nu) = \phi(\lambda_{\theta_0, r}(W, \mathcal{N}, \nu), a, b)$$

for any  $\nu$ .

The following proposition is immediate from Lemma 1, Assumption 13, and Assumption 4(b).

**Proposition 5.** *For any  $\mathcal{N} \subseteq \mathbb{R}^d$ , suppose the network  $G(\cdot, \cdot, \mathcal{N}, r)$  is an equilibrium network for some  $r$ . Then under Assumptions 12 and 13,  $G$  is stationary.*

A sufficient condition for Assumption 12 is the *stationarity condition* that  $S$  and  $W$  depend on  $\mathcal{N}$  only through  $\{r^{-1}\|i - j\|; i, j \in \mathcal{N}\}$ . The latter holds if  $W_{ij} = (r^{-1}\|i - j\|, Z_i, Z_j, \zeta_{ij})$ , and  $\zeta_{ij}$  depends on  $\mathcal{N}$  only through  $r^{-1}\|i - j\|$ . While it is enough to assume that  $S$  does not depend directly on  $\mathcal{N}$  to Assumption 12(b) to hold, the same condition for  $\lambda$  is not quite sufficient for Assumption 13. An additional restriction is still needed to ensure that  $\lambda_{\theta_0, ar}(W, a\mathcal{N} + b, \nu)$  is mapping to the analogous isomorphic equilibrium in  $\mathcal{G}_{\theta_0, r}(W, \mathcal{N})$ , hence Assumption 13.

## A.2 A General Weak Law

Let  $\xi(i, \mathcal{S}, h)$  be an  $\mathbb{R}$ -valued function defined for all locally finite  $\mathcal{S} \subseteq \mathbb{R}^d$ ,  $i \in \mathcal{S}$ , and functions  $h(\cdot, \cdot, \mathcal{S}) : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^{d_w}$ . For ease of notation, if  $i \notin \mathcal{N}$ , we let  $\xi(i, \mathcal{N}, h) \equiv \xi(i, \mathcal{N} \cup \{i\}, h)$ .

For any  $\mathcal{S} \subseteq \mathbb{R}^d$ , define the random function  $\phi(\mathcal{S}) \equiv \phi(\cdot, \cdot, \mathcal{S}) : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^{d_w}$  such that there exist independently distributed  $d_z$ -dimensional random vectors  $\{Z_i; i \in \mathcal{S}\}$  and independently distributed  $d_\zeta$ -dimensional random vectors  $\{\zeta_{ij}; i, j \in \mathcal{S}\}$ , both independent of  $\mathcal{N}$ , for which  $\phi(i, j) = (Z_i, Z_j, \zeta_{ij})$ . The

We next adapt the notion of strongly stabilizing functionals in [Penrose and Yukich \(2001\)](#) to our context.<sup>24</sup>

**Definition 3.** The functional  $\xi$  is *strongly stabilizing* on a locally finite set  $\mathcal{B} \subseteq \mathbb{R}^d$  if for any  $i \in \mathbb{R}^d$ , there exists a radius  $R < \infty$  such that for all locally finite  $\mathcal{A} \subseteq \mathbb{R}^d \setminus B(i, R)$ ,

$$\xi(i, \mathcal{B} \cup \{i\}, \phi) = \xi(i, (\mathcal{B} \cup \{i\}) \cap B(i, R) \cup \mathcal{A}, \omega)$$

with probability one, for some  $\phi$ -dependent function  $\omega : ((\mathcal{B} \cup \{i\}) \cap B(i, R) \cup \mathcal{A})^2 \rightarrow \mathbb{R}^{d_w}$  satisfying  $\omega(j, k) = \phi(j, k, \mathcal{B} \cup \{i\})$  whenever  $j, k \in \mathcal{B} \cup \{i\}$ .

The following theorem adapts Theorem 2.1 of [Penrose and Yukich \(2003\)](#) (PY) to strongly stabilizing functionals  $\xi$  that depend on  $\phi$ .

**Theorem 3.** *Suppose  $\xi$  is strongly stabilizing on  $\mathcal{P}_\tau$  for any  $\tau$  in the range of  $\kappa f(\cdot)$ . If*

$$\sup_{n \in \mathbb{N}} \mathbf{E} [\xi(i, i + r_n^{-1}(\mathcal{N} - i), \phi)^p] < \infty$$

for some  $p > 2$  and  $nr_n^d \rightarrow \kappa$ , then

$$\frac{1}{n} \sum_{i \in \mathcal{N}_n} \xi(i, i + r_n^{-1}(\mathcal{N}_n - i), \phi) \xrightarrow{L_2} \int \mathbf{E}[\xi(i, \mathcal{P}_{\kappa f(i)}, \phi) \mid i] f(i) di.$$

**Coupling  $\mathcal{N}$ .** To prove the theorem, we couple the processes  $i + r_n^{-1}(\mathcal{N}_n - i)$  and  $\mathcal{P}_{\kappa f(i)}$  using a construction due to PY. Let  $i$  be a random variable distributed with density  $f$  and  $\mathcal{P}_1$  be a Poisson process of rate one on  $\mathbb{R}^d \times [0, \infty)$  independent of  $i$ . Let  $\mathcal{P}_{nf}^*$  be the restriction of  $\mathcal{P}_1$  to  $\{(k, t) \in \mathbb{R}^d \times [0, \infty) : t \leq nf(k)\}$  and  $\mathcal{P}_{nf}$  the image of  $\mathcal{P}_{nf}^*$  under the projection  $(k, t) \mapsto k$ .<sup>25</sup> Let  $N(\mathcal{P}_{nf})$  be the number of points in  $\mathcal{P}_{nf}$ . Construct  $\mathcal{N}'_{n-1}$  from  $N(\mathcal{P}_{nf})$  by dropping  $(N(\mathcal{P}_{nf}) - (n-1))^+$  points randomly and including  $(n-1 - N(\mathcal{P}_{nf}))^+$  independent points drawn from  $f$ .<sup>26</sup> Let  $\mathcal{P}_{nf(i)}^*$  be the restriction of  $\mathcal{P}_1$  to  $\{(k, t) : t \leq nf(i)\}$  and  $\mathcal{P}_{\kappa f(i)}$  the image of  $\mathcal{P}_{nf(i)}^*$  under  $(k, t) \mapsto (\frac{n}{\kappa})^d(k - i)$ . Then conditional on  $i$ ,  $\mathcal{P}_{\kappa f(i)}$  is a homogeneous Poisson process on  $\mathbb{R}^d$  of intensity  $\kappa f(i)$  by the ‘‘mapping theorem’’ for Poisson processes (e.g. [Kingman, 1992](#)).

<sup>24</sup>An analogous extension of the weaker notion of stabilization in [Penrose and Yukich \(2003\)](#) can also be defined. Strong stabilization suffices for our purposes.

<sup>25</sup>Then  $\mathcal{P}_{nf}$  is an inhomogeneous Poisson point process with intensity  $nf(\cdot)$ .

<sup>26</sup>Then  $\mathcal{N}'_{n-1}$  and  $\mathcal{N}_{n-1}$  are identically distributed.

## A WEAK LAW FOR NETWORK MOMENTS

**Coupling**  $\phi(\mathcal{N})$ . For any  $i \in \mathbb{R}^d$ , we couple  $\phi(i + r_n^{-1}(\mathcal{N}'_{n-1} - i) \cup \{i\})$  to  $\phi(\mathcal{P}_{\kappa f(i)} \cup \{i\})$  as follows. For  $k \in \mathbb{R}^d$ ,  $t \in \mathbb{R}$ , define  $\varrho : (k, t) \mapsto k$ . Let  $\{Z_j; j \in \mathbb{R}^d\}$  and  $\{\zeta_{jk}; j, k \in \mathbb{R}^d\}$  be independently distributed random vectors of dimension  $d_z$  and  $d_\zeta$ , respectively. In what follows, let  $i' = (i, 0)$ .

- For  $k \in \mathbb{R}^d$ ,  $t \in \mathbb{R}$ , define  $\varrho_i : (k, t) \mapsto \left(\frac{n}{\kappa}\right)^d (k - i)$ . For each  $j, k \in \mathcal{P}_{nf(i)}^* \cup \{i'\}$ , let

$$\phi(\varrho_i(j), \varrho_i(k), \mathcal{P}_{\kappa f(i)} \cup \{i\}) = (Z_{\varrho_i(j)}, Z_{\varrho_i(k)}, \zeta_{\varrho_i(j)\varrho_i(k)}).$$

- For each  $j, k \in \mathcal{P}_{nf}^* \cup \{i'\}$ , define

$$\phi(\varrho(j), \varrho(k), \mathcal{P}_{nf} \cup \{i\}) = (Z_{\varrho(j)}, Z_{\varrho(k)}, \zeta_{\varrho(j)\varrho(k)}).$$

For  $k \in \mathbb{R}^d$ ,  $t \in \mathbb{R}$ , define  $\varrho'_i : (k, t) \mapsto i + r_n^{-1}(k - i)$ . For each  $j, k \in \mathcal{P}_{nf}^* \cup \{i'\}$  for which there exist  $j', k' \in \mathcal{N}'_{n-1} \cup \{i\}$  such that  $j' = \varrho(j)$ ,  $k' = \varrho(k)$ , define

$$\phi(\varrho'_i(j), \varrho'_i(k), i + r_n^{-1}(\mathcal{N}'_{n-1} - i) \cup \{i\}) = (Z_{\varrho(j)}, Z_{\varrho(k)}, \zeta_{\varrho(j)\varrho(k)}).$$

**Lemma 2** (Coupling). *For any  $M > 0$ ,*

$$\lim_{n \rightarrow \infty} \mathbf{P}(r_n^{-1}(\mathcal{N}'_{n-1} - i) \cap B(i, M) = \mathcal{P}_{\kappa f(i)} \cap B(i, M)) = 1.$$

PROOF. This follows from the proof of Lemma 3.1 of PY. We give a fully-detailed argument here for completeness.

By the Lebesgue density theorem, with probability one,  $i$  is realized at a Lebesgue point  $f$ . Conditional on  $i$ , the expected number of points  $(j, t)$  of  $\mathcal{P}_1$  in  $B(i, r_n M) \times [0, \infty)$  that satisfy  $t \leq r_n^{-d} \kappa f(i)$  (and hence contribute to  $\mathcal{P}_{\kappa f(i)}$ ) but also satisfy  $t > nf(j)$  (and hence do not contribute to  $\mathcal{P}_{nf}^*$ ) is

$$\begin{aligned} \int_{B(i, r_n M)} \int_{r_n^{-d} \kappa f(j)}^{nf(i)} \mathbf{1}\{nf(i) > nf(j)\} dt dj &= \int_{B(i, r_n M)} (r_n^{-d} \kappa f(i) - nf(j))^+ dj \\ &\leq r_n^{-d} \int_{B(i, r_n M)} \kappa |f(i) - f(j)| dj + r_n^{-d} \int_{B(i, r_n M)} |\kappa - nr_n^d| f(j) dj, \end{aligned}$$

where the left-hand side of the first line is due to the ‘‘mapping theorem’’ for Poisson processes (e.g. Kingman, 1992). The last line converges to zero since  $i$  is a Lebesgue point of  $f$  and  $nr_n^d \rightarrow \kappa$ . Similarly, the expected number of points of  $\mathcal{P}_1$  in  $B(i, r_n M) \times [0, \infty)$  that contribute to  $\mathcal{P}_{nf}^*$  but not to  $\mathcal{P}_{\kappa f(i)}$  also tends to zero. Therefore, the probability that  $r_n^{-1} \mathcal{P}_{nf}^* \cap B(i, M)$  and  $\mathcal{P}_{\kappa f(i)} \cap B(i, M)$  are equivalent tends to one.

Let  $E = \{\mathcal{P}_{nf}^* \cap B(i, r_n M) \neq \mathcal{N}'_{n-1} \cap B(i, r_n M)\}$ . We next show that  $\mathbf{P}(E) \rightarrow 0$ . Notice that  $E$  occurs either if any of the  $(N(P_1^n) - n - 1)^+$  discarded points or the  $(n - 1 - N(P_1^n))^+$  added points in the construction of  $\mathcal{N}'_{n-1}$  lies in  $B(i, r_n M)$ . Denote this event by  $E'$ . For any  $\epsilon > 0$ , choose  $\tilde{n}$  large enough such that (a) for some fixed  $\bar{\kappa}$ ,  $\tilde{n} r_{\tilde{n}}^d \leq \bar{\kappa}$  for all  $n > \tilde{n}$ , and (b),

$$\mathbf{P}\left(|N(\mathcal{P}_{nf}^*) - n - 1| > \frac{n\epsilon}{2c\bar{\kappa}}\right) < \left(1 - \frac{\kappa}{\bar{\kappa}}\right) \epsilon, \quad (34)$$

where  $c$  satisfies  $\mathbf{P}(\|i - j\| \leq r_n M \mid i) \leq cr_n^d$  for  $j \sim f(\cdot)$ . Such a  $c$  exists by Assumption 6. Such a  $\tilde{n}$  exists because

$$\begin{aligned} \mathbf{P}(|N(\mathcal{P}_{n_f}^*) - n - 1| > c'n) &\leq \mathbf{P}((N(\mathcal{P}_{n_f}^*) - n)^2 > (c'n - 1)^2) \\ &\leq \frac{\text{Var}(N(\mathcal{P}_{n_f}^*))}{(c'n - 1)^2} = \frac{n}{(c'n - 1)^2}, \end{aligned}$$

so  $|N(\mathcal{P}_{n_f}^*) - n - 1| = o_p(n)$ . Therefore,

$$\begin{aligned} \mathbf{P}(E) &\leq \mathbf{P}\left(E' \mid |N(\mathcal{P}_{n_f}^*) - n - 1| \leq \frac{n\epsilon}{c\bar{\kappa}}\right) + \mathbf{P}\left(|N(\mathcal{P}_{n_f}^*) - n - 1| > \frac{n\epsilon}{c\bar{\kappa}}\right) \\ &\leq \frac{n\epsilon}{c\bar{\kappa}} cr_n^d + \left(1 - \frac{\kappa}{\bar{\kappa}}\right) \epsilon \leq \epsilon, \end{aligned}$$

which proves the claim that  $\mathbf{P}(E) \rightarrow 0$ .

The above arguments establish that

$$\lim_{n \rightarrow \infty} \mathbf{P}\left(r_n^{-1}(\mathcal{N}'_{n-1} - i) \cap B(i, M) = \mathcal{P}_{\kappa f(i)} \cap B(i, M) \mid i\right) = 1$$

almost surely. The result follows by dominated convergence.  $\blacksquare$

For  $i \in \mathbb{R}^d$ , define  $\varsigma(i) = \xi(i, \mathcal{P}_{\kappa f(i)}, \phi)$ , and let  $B(i, m)$  be the ball of radius  $m$  centered at  $i$ . For  $\mathcal{B} \subseteq \mathbb{R}^d$  locally finite and  $m \in \mathbb{R}^+$ , define

$$\begin{aligned} \bar{\xi}(i, \mathcal{B}, \phi, m) &= \sup_{\ell \in \mathbb{N}} \text{ess sup}_{|\mathcal{A}|=\ell} \xi(i, (\mathcal{B} \cup \{i\}) \cap B(i, m) \cup \mathcal{A}, \omega), \\ \underline{\xi}(i, \mathcal{B}, \phi, m) &= \inf_{\ell \in \mathbb{N}} \text{ess inf}_{|\mathcal{A}|=\ell} \xi(i, (\mathcal{B} \cup \{i\}) \cap B(i, m) \cup \mathcal{A}, \omega). \end{aligned}$$

where the essential supremum and infimum are taken with respect to the Lebesgue measure over the set of  $\mathcal{A} \subseteq \mathbb{R}^d \setminus B(i, m)$  with cardinality  $\ell$ .

**Lemma 3** (Convergence of Means). *Under the assumptions of Theorem 3,*

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[ \frac{1}{n} \sum_{i \in \mathcal{N}_n} \xi(i, i + r_n^{-1}(\mathcal{N}_n - i), \phi) \right] = \int \mathbf{E} [\xi(i, \mathcal{P}_{\kappa f(i)}, \phi)] f(i) di. \quad (35)$$

PROOF. We follow the proof of Lemma 3.2 of PY. For any  $\varepsilon, M > 0$ ,

$$\begin{aligned} \mathbf{P}\left(|\xi(i, i + r_n^{-1}(\mathcal{N}'_{n-1} - i), \phi) - \varsigma(i)| > \varepsilon\right) \\ \leq \mathbf{P}\left(i + r_n^{-1}(\mathcal{N}'_{n-1} - i) \cap B(i, M) \neq \mathcal{P}_{\kappa f(i)} \cap B(i, M)\right) \\ + \mathbf{P}\left(\bar{\xi}(i, \mathcal{P}_{\kappa f(i)}, \phi, M) - \underline{\xi}(i, \mathcal{P}_{\kappa f(i)}, \phi, M) > \varepsilon\right). \end{aligned}$$

The first part of the right-hand side converges to zero by Lemma 2. The second part is zero for  $M$  sufficiently large, since  $\xi$  is strongly stabilizing. Therefore,

$$\xi(i, i + r_n^{-1}(\mathcal{N}'_{n-1} - i), \phi) \xrightarrow{d} \varsigma(i). \quad (36)$$

## A WEAK LAW FOR NETWORK MOMENTS

By the almost-sure representation theorem and uniform integrability (10),

$$\mathbf{E} [\xi (i, i + r_n^{-1}(\mathcal{N}'_{n-1} - i), \phi)] \rightarrow \mathbf{E}[\zeta(i)].$$

Now, by the law of iterated expectations, conditioning on  $i$ , the right-hand side equals the right-hand side of (35). Notice that the left-hand side equals

$$\mathbf{E} \left[ \frac{1}{n} \sum_{i \in \mathcal{N}_n} \xi (i, i + r_n^{-1}(\mathcal{N}_n - i), \phi) \right]$$

by identical distribution. This completes the proof. ■

To prove Theorem 3, we need to extend the previous coupling construction to establish concentration of the variance.

**Coupling  $\mathcal{N}$ .** Let  $i$  and  $j$  be independently drawn from  $f$ , and let  $\mathcal{P}_1, \mathcal{Q}_1$  be independent Poisson processes on  $\mathbb{R}^d \times [0, \infty)$ . Derive  $\mathcal{P}_{nf}$  from  $\mathcal{P}_1$  as before and similarly derive  $\mathcal{Q}_{nf}$  from  $\mathcal{Q}_1$ . Construct  $\mathcal{N}'_{n-2}$  from  $N(\mathcal{P}_{nf})$  by dropping  $(N(\mathcal{P}_{nf}) - (n-2))^+$  points randomly and including  $(n-2 - N(\mathcal{P}_{nf}))^+$  independent points drawn from  $f$ . Let  $F_i$  be the half-space of points in  $\mathbb{R}^d$  closer to  $i$  than to  $j$  and  $F_j$  the points closer to  $j$ . Let  $\mathcal{P}_{nf(i)}^i$  be the restriction of  $\mathcal{P}_1$  to  $F_i \times [0, nf(i)]$  and  $\mathcal{Q}_{nf(i)}^j$  the restriction of  $\mathcal{Q}_1$  to  $F_j \times [0, nf(i)]$ . Let  $\mathcal{P}_{\kappa f(i)}^+$  be the image of  $\mathcal{P}_{nf(i)}^i \cup \mathcal{Q}_{nf(i)}^j$  under  $(k, t) \mapsto \left(\frac{n}{\kappa}\right)^d (k - i)$ .

Analogously, let  $\mathcal{P}_{nf(j)}^j$  be the restriction of  $\mathcal{P}_1$  to  $F_j \times [0, nf(j)]$  and  $\mathcal{Q}_{nf(j)}^i$  the restriction of  $\mathcal{Q}_1$  to  $F_i \times [0, nf(j)]$ . Let  $\mathcal{P}_{\kappa f(j)}^+$  be the image of  $\mathcal{P}_{nf(j)}^j \cup \mathcal{Q}_{nf(j)}^i$  under  $(k, t) \mapsto \left(\frac{n}{\kappa}\right)^d (k - j)$ . Then

$$\mathcal{P}_{\kappa f(i)}^+ \perp \mathcal{P}_{\kappa f(j)}^+ \tag{37}$$

by the spatial independence property of Poisson processes.

**Coupling  $\phi(\mathcal{N})$ .** Let  $\phi', \phi'', \phi'''$  be independent copies of  $\phi$ , where  $\phi$  is defined similarly to the previous coupling construction.<sup>27</sup> For ease of notation, define  $\mathcal{N}(i, j) = (i + r_n^{-1}(\mathcal{N}'_{n-2} \cup \{j\} - i)) \cup \{i\}$ . Let  $\phi_i^+ : \mathbb{R}^{2d} \rightarrow \{0, 1\}$  satisfy

$$\phi_i^+(k, l, \mathcal{N}(i, j)) = \begin{cases} \phi(k, l, \mathcal{N}(i, j)) & \text{if } k, l \in F_i \\ \phi'(k, l, \mathcal{N}(i, j)) & \text{if } k, l \in F_j \\ \phi''(k, l, \mathcal{N}(i, j)) & \text{otherwise,} \end{cases}$$

$$\phi_i^+(k, l, \mathcal{P}_{\kappa f(i)} \cup \{i\}) = \begin{cases} \phi(k, l, \mathcal{P}_{\kappa f(i)} \cup \{i\}) & \text{if } k, l \in F_i \\ \phi'(k, l, \mathcal{P}_{\kappa f(i)} \cup \{i\}) & \text{if } k, l \in F_j \\ \phi''(k, l, \mathcal{P}_{\kappa f(i)} \cup \{i\}) & \text{otherwise,} \end{cases}$$

---

<sup>27</sup>In the definition, replace  $\mathcal{N}'_{n-1}$  with  $\mathcal{N}'_{n-2}$ , and replace  $\{i\}$  with  $\{i, j\}$ .

Similarly let  $\phi_j^+ : \mathbb{R}^{2d} \rightarrow \{0, 1\}$  satisfy

$$\begin{aligned} \phi_j^+(k, l, \mathcal{N}(j, i)) &= \begin{cases} \phi'(k, l, \mathcal{N}(j, i)) & \text{if } k, l \in F_i \\ \phi(k, l, \mathcal{N}(j, i)) & \text{if } k, l \in F_j \\ \phi'''(k, l, \mathcal{N}(j, i)) & \text{otherwise,} \end{cases} \\ \phi_j^+(k, l, \mathcal{P}_{\kappa f(j)} \cup \{j\}) &= \begin{cases} \phi'(k, l, \mathcal{P}_{\kappa f(j)} \cup \{j\}) & \text{if } k, l \in F_i \\ \phi(k, l, \mathcal{P}_{\kappa f(j)} \cup \{j\}) & \text{if } k, l \in F_j \\ \phi'''(k, l, \mathcal{P}_{\kappa f(j)} \cup \{j\}) & \text{otherwise,} \end{cases} \end{aligned}$$

Note that  $\phi$  does not have the same distribution as  $\phi_i^+$  or  $\phi_j^+$ . Lastly, for  $x \in \{i, j\}$  define  $\varsigma^+(x) = \xi(x, \mathcal{P}_{\kappa f(x)}^+, \phi_x^+)$ . Then by (37) and construction of  $\phi, \phi', \phi'', \phi'''$ ,

$$\varsigma^+(i) \perp \varsigma^+(j). \quad (38)$$

PROOF OF THEOREM 3. Given Lemma 3, it remains to show concentration, that the variance converges to zero. Notice that for any  $M > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbf{P}(B(i, Mr_n) \subseteq F_i) = \lim_{n \rightarrow \infty} \mathbf{P}(B(j, Mr_n) \subseteq F_j) = 1.$$

This and an argument similar to the proof of Lemma 2 imply that for any  $M > 0$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}\left(i + r_n^{-1}((\mathcal{N}'_{n-2} \cup j) - i) \cap B(i, M) = \mathcal{P}_{\kappa f(i)}^+ \cap B(i, M)\right) &= 1, \\ \lim_{n \rightarrow \infty} \mathbf{P}\left(j + r_n^{-1}((\mathcal{N}'_{n-2} \cup i) - j) \cap B(j, M) = \mathcal{P}_{\kappa f(j)}^+ \cap B(j, M)\right) &= 1. \end{aligned}$$

Using these equations and equality of  $\phi$  and  $\phi_i^+$  on  $F_i$ , we have for any  $\varepsilon, M > 0$ ,

$$\begin{aligned} &\mathbf{P}\left(|\xi(i, i + r_n^{-1}((\mathcal{N}'_{n-2} \cup \{j\}) - i), \phi) - \varsigma^+(i)| > \varepsilon\right) \\ &\leq \mathbf{P}\left(i + r_n^{-1}((\mathcal{N}'_{n-2} \cup \{j\}) - i) \cap B(i, M) \neq \mathcal{P}_{\kappa f(i)}^+ \cap B(i, M)\right) \\ &+ \mathbf{P}\left(\phi\left((i + r_n^{-1}(\mathcal{N}'_{n-2} \cup \{j\}) - i) \cup \{i\}\right) \Big|_{B(i, M)} \neq \phi_i^+\left(\mathcal{P}_{\kappa f(i)}^+ \cup \{i\}\right) \Big|_{B(i, M)}\right) \\ &+ \mathbf{P}\left(\bar{\xi}\left(i, \mathcal{P}_{\kappa f(i)}^+, \phi_i^+, M\right) - \xi\left(i, \mathcal{P}_{\kappa f(i)}^+, \phi_i^+, M\right) > \varepsilon\right), \end{aligned}$$

where for any  $B \subseteq \mathbb{R}^d$ ,  $\phi|_B$  is the restriction of  $\phi$  to the set  $B \times B$ . The right-hand side tends to zero as  $n \rightarrow \infty$  for  $M$  sufficiently large, using Lemma 2 and strong stability of  $\xi$ . Thus, as in (36),

$$\xi(i, i + r_n^{-1}((\mathcal{N}'_{n-2} \cup \{j\}) - i), \phi) \times \xi(j, j + r_n^{-1}((\mathcal{N}'_{n-2} \cup \{i\}) - j), \phi) \xrightarrow{d} \varsigma^+(i)\varsigma^+(j).$$

By the almost-sure representation theorem, uniform square-integrability (10), identical distribution of  $\mathcal{N}'_{n-2} \cup \{i, j\}$  and  $\mathcal{N}_n$ , identical distribution of  $\phi, \phi_i^+, \phi_j^+$ , and (38),

$$\lim_{n \rightarrow \infty} \mathbf{E}\left[\xi(i, i + r_n^{-1}(\mathcal{N}_n - i), \phi) \xi(j, j + r_n^{-1}(\mathcal{N}_n - j), \phi)\right] = \mu^2. \quad (39)$$

Finally, notice that

$$\begin{aligned} \mathbf{E} \left[ \left( \frac{1}{n} \sum_{i \in \mathcal{N}_n} \xi(i, i + r_n^{-1}(\mathcal{N}_n - i), \phi) \right)^2 \right] &= \frac{1}{n} \mathbf{E} [\xi(i, i + r_n^{-1}(\mathcal{N}_n - i), \phi)^2] \\ &\quad + \left( 1 - \frac{1}{n} \right) \mathbf{E} [\xi(i, i + r_n^{-1}(\mathcal{N}_n - i), \phi) \xi(j, j + r_n^{-1}(\mathcal{N}_n - j), \phi)]. \end{aligned}$$

This tends to  $\mu^2$  by uniform integrability of  $\xi$  and (39), and thus the variance tends to zero.  $\blacksquare$

### A.3 Proofs of Main Results

PROOF OF PROPOSITION 1. Fix  $\alpha$  in the support of  $\kappa f(\cdot)$ , and set  $\mathcal{N} = \mathcal{P}_\alpha$ .

Consider a breadth-first search (BFS) of  $D(1)$  starting at an arbitrary node  $i$ . Initialize a queue (ordered set) of “saturated” nodes  $S_0 = \{i\}$  and a set of “explored” nodes  $E_0 = \emptyset$ . At step  $m$ , if  $S_{m-1} \neq \emptyset$ , pop (remove) a node  $j$  from the top of the queue  $S_{m-1}$ , and define  $S_m = S_{m-1} \setminus \{j\}$  and  $E_m = E_{m-1} \cup \{j\}$ . Then push (add to the end) each node in  $\{k : D_{jk}(1) = 1, k \notin S_{m-1}\}$ , the set of neighbors of  $j$  not already queued, into the end of the queue  $S_m$  in some arbitrary order. Let  $O_m^*$  denote the number of neighbors pushed in the  $m$ th step.

We next construct a process  $\{O_m\}$  using a backtracking breadth-first search (BBFS). Again, let  $S_0 = \{i\}$  and  $E_0 = \emptyset$ . At step  $m$ , if  $S_{m-1} \neq \emptyset$ , pop a node  $j$  from  $S_{m-1}$ , and define  $S_m = S_{m-1} \setminus \{j\}$  and  $E_m = E_{m-1} \cup \{j\}$ . Then push each node in  $\{k : D_{jk}(r_n) = 1\}$ , the set of *all* neighbors of  $j$ , into the end of the queue  $S_m$  in some arbitrary order. Let  $O_m^*$  denote the number of neighbors pushed at the  $m$ th step.

The difference between the two searches is that the BBFS backtracks to previously searched nodes, while the BFS does not, so  $\sum_{m=1}^\infty O_m \geq \sum_{m=1}^\infty O_m^*$ . It therefore suffices to show that the process  $\{O_m\}$  dies in finite time ( $\lim_{m \rightarrow \infty} \mathbf{P}(O_m = 0) = 1$ ).

Note that  $\{O_m\}$  is a multi-type Galton-Walton branching process in which the offspring distribution of a type  $(i, Z_i = z)$  individual is the inhomogeneous Poisson point process on  $\mathbb{R}^d$  with intensity function

$$\tau_{i,z}(j, z') = \alpha \mathbf{P}(E_{ij}^c(1) \cap \{\Pi_{ij}(1) = 1\} \mid i, j, Z_i = z, Z_j = z') f_z(z'),$$

where  $f_z(\cdot)$  is the density of  $Z_j$  for any fixed  $j$ . The expected number of offspring is  $\int \int \tau_{i,z}(j, z') d_{z'} dj$ , which is the conditional expected degree of a node of type  $i$  with  $Z_i = z$  for the graph  $D(1)$  when  $\mathcal{N} = \mathcal{P}_\alpha$ , as expected.

Let  $\rho(i, z)$  be the probability that the process  $\{O_m\}$  survives indefinitely ( $\lim_{m \rightarrow \infty} \mathbf{P}(O_m = 0) = 0$ ) when the type of the starting individual is  $(i, z)$ . Let  $\mathcal{H}$  be the class of functions  $h : (i, z) \mapsto [0, 1]$ ,

$$\begin{aligned} k : (i, j, Z_i, Z_j) &\mapsto \alpha \mathbf{P}(E_{ij}^c(1) \cap \{\Pi_{ij}(1) = 1\} \mid i, j, Z_i = z, Z_j = z') f(j)^{-1}, \text{ and} \\ T : h &\mapsto \mathbf{E}[k(i, j, Z_i, Z_j) h(j, Z_j)]. \end{aligned}$$

By Lemma 5.11 of Bollobás et al. (2007), since  $\sup_{h \in \mathcal{H}} \|Th\|_2 \leq 1$  (Assumption 7),<sup>28</sup> then  $\rho(i, Z_i) = 0$  almost surely ( $\|\cdot\|_q$  denotes the  $L_q$  norm), so the branching process  $\{O_m\}$  dies out in finite time, and the proof is complete.  $\blacksquare$

**PROOF OF PROPOSITION 2.** Consider the BFS detailed in Proposition 1 but applied to  $D(r_n)$  rather than  $D(1)$  and the induced branching process  $\{O_m^*\}$ . Let  $\{O'_m\}$  be i.i.d.  $\text{Bin}(n, c_n)$  random variables, where  $c_n = \min\{r_n^d \gamma_n^* \int \text{ess sup}_i p_{ij}(r_n) dj, 1\}$ . Then  $\{O'_m\}$  stochastically dominates  $\{O_m^*\}$ .

An arbitrary node  $i$  lies in a component of  $D(r_n)$  containing at least  $M$  nodes only if the BFS starting at  $i$  finds at least  $M - 1$  explored nodes after  $M - 1$  BFS steps. By stochastic dominance, then, the probability of the latter event is

$$\begin{aligned} \mathbf{P}\left(\sum_{m=1}^{M-1} O_m^* \geq M - 1\right) &\leq \mathbf{P}\left(\sum_{m=1}^{M-1} O'_m \geq M - 1\right) \\ &= \mathbf{P}\left(\sum_{m=1}^{M-1} O'_m \geq nc_n(M - 1) \left(1 + \frac{(1 - nc_n)(M - 1) - 2}{nc_n(M - 1)}\right)\right) \\ &\leq \exp\left\{-\frac{(M - nc_n(M - 1) - 1)^2}{M + nc_n(M - 1) - 1}\right\}, \end{aligned}$$

where the last line uses the Chernoff bound. Since  $nc_n$  tends to a constant that is strictly less than one by Assumptions 6 and 7, if we set  $M = \frac{b}{(1 - nc_n)^2} \log n$  for  $b > 0$ , then for  $n$  sufficiently large, the last line is bounded by

$$\exp\left\{-\frac{M^2(1 - nc_n)^2}{2M} + (1 - nc_n)^2\right\},$$

which is  $O(n^{-b/2})$ . Finally, by the union bound, the probability that there exists a node  $i$  for which  $\mathbf{P}(|C_i(D)| \geq M)$  is  $O(n^{1-b/2}) \rightarrow 0$  for  $b$  large.  $\blacksquare$

The following lemma is used in the proof of Theorem 1.

**Lemma 4.** *For any  $\tau$  in the support of  $\kappa f(\cdot)$ ,  $\xi(i, \mathcal{P}_\tau, W) \equiv \mathbf{E}[\psi_i(\mathcal{P}_\tau, 1) | W, \mathcal{P}_\tau \cup \{i\}]$  is strongly stabilizing on  $\mathcal{P}_\tau$  under Assumptions 6 and 7.*

**PROOF.** By Proposition 1 and Assumption 6,  $|J_i|$  is almost surely finite. Thus with probability one, there exists a radius  $R < \infty$  such that  $B(i, R)$  encompasses all nodes in  $J_i$ . By locality of  $\psi$  (Definition 1) and Assumptions 2 and 5,  $\mathbf{E}[\psi_i(\mathcal{P}_\tau, 1) | W, \mathcal{P}_\tau \cup \{i\}]$  is invariant to changes in  $\mathcal{N}$  and  $W$  so long as  $J_i$  and  $\{W_{jk}; j, k \in J_i\}$  remain the same. This motivates next our construction of  $r$  to satisfy Definition 3. Using Assumption 1, for any  $j, k \in \mathcal{P}_\tau$ , let  $z_e(Z_j), \zeta_e(Z_j), z_a(Z_j), \zeta_a(Z_j)$  satisfy

$$\max\{V(s, (Z_j, z_e(Z_j), \zeta_e(Z_j)); \theta_0), V(s, (z_a(Z_k), Z_k, \zeta_a(Z_k)); \theta_0)\} \leq 0.$$

<sup>28</sup>They take  $\mathcal{H}$  to be the class of functions  $h : (i, z) \mapsto \mathbb{R}^+$  with  $\|h\|_2 \leq 1$ , but it is sufficient to consider  $h$  such that  $0 \leq h \leq 1$ , since solutions of their equation (5.3) necessarily obey this condition by their Lemma 5.8(i).

## A WEAK LAW FOR NETWORK MOMENTS

Then define

$$\omega(j, k) = \begin{cases} (Z_i, Z_j, \zeta_{ij}) & \text{if } j, k \in \mathcal{P}_\tau \\ (Z_i, Z_j, \zeta_{ij}) & \text{if } j, k \in \mathcal{A} \\ (Z_i, z_e(Z_i), \zeta_e(Z_i)) & \text{if } j \in \mathcal{P}_\tau, k \in \mathcal{A} \\ (z_a(Z_k), Z_k, \zeta_a(Z_k)) & \text{if } j \in \mathcal{A}, k \in \mathcal{P}_\tau \end{cases}$$

It follows that

$$\mathbf{E}[\psi_i(\mathcal{P}_\tau, 1) | \omega, (\mathcal{P}_\tau \cup \{i\}) \cap B(i, R) \cup \mathcal{A} \cup] = \mathbf{E}[\psi_i(\mathcal{P}_\tau, 1) | W, \mathcal{P}_\tau \cup \{i\}]$$

for any  $\mathcal{A} \subseteq \mathbb{R}^d \setminus B(i, R)$ , with probability one. This is because, by construction of  $r$ , no node pair  $j \in \mathcal{P}_\tau \cap B(i, R), k \in \mathcal{A}$  enjoys positive surplus from forming a link, so  $J_i$  is preserved. This completes the proof.  $\blacksquare$

**PROOF OF THEOREM 1.** Recall the definitions of [II] and [III] from (13).  $L_2$  convergence of [II] follows from condition (11), Lemma 4, and Theorem 3. Turning to [III], by an argument similar to Lemma 4,  $|J_i|$  can be written as a stabilizing functional  $\xi(i, \mathcal{P}_\tau, W)$  and therefore converges to a limit. As argued in the proof sketch of this theorem in §5.1, this limit is finite. Convergence of [I] then follows by dominated convergence.  $\blacksquare$

**PROOF OF THEOREM 2.** We first prove the theorem holds if we replace  $\mathcal{U}$  with  $\tilde{\mathcal{U}} = \{u \in \mathbb{R}^{|\mathcal{Y}|} : \|u\| \leq 1\}$  by verifying the conditions of Theorem 2.1 in Beresteanu et al. (2011). Then the result follows from Theorem D.1 of Beresteanu et al. (2011).

Assumption 9 implies BMM's Assumption 2.1, and their Assumptions 2.2-2.3 hold by construction of  $Q_\theta$ . Turning to their Assumption 2.4, for  $y \in \mathcal{Y}$ , define selection mechanisms for stable dyadic outcomes (henceforth dyadic SMs),  $\sigma_{\theta_0, r_n}(y | W_{ij})$ , to be conditional distributions with support  $\mathcal{S}_{\theta_0}(W_{ij}, r_n)$ . Define one such dyadic SM induced by a selection mechanism for stable networks  $\lambda$ :

$$\sigma_{\theta_0, r_n}^*(y | w) = \mathbf{E} \left[ \sum_{g \in \mathbf{G}_{ij}(y; W)} \mathbf{P}(\lambda_{\theta_0, r_n}(W, \mathcal{N}_n, \nu) = g | W, \mathcal{N}_n) \Big| W_{ij} = w \right], \quad (40)$$

where  $\mathbf{G}_{ij}(y; W) \subseteq \mathbf{G}_n$  is the set of networks  $g$  such that  $(g_{ij}, S_{ij}(g, W)) = y$  under  $W$ . The support of this distribution is  $\mathcal{S}_{\theta_0}(W_{ij}, r_n)$ , as claimed, by Assumption 4(b). This establishes their Assumption 2.4.

We lastly verify their Assumption 2.5, that our Assumption 4(b) holds if and only if there exists a dyadic SM  $\sigma_{\theta_0, r}$  rationalizing the distribution of dyadic outcomes in the sense that for any  $y \in \mathcal{Y}$ ,

$$\mathbf{P}(Y_{ij} = y | X_{ij}) = \mathbf{E}[\sigma_{\theta_0, r_n}(y | W_{ij}) | X_{ij}].$$

The ‘‘only if’’ direction holds by construction (40). To see the ‘‘if’’ direction, suppose that Assumption 4(b) fails to hold. Then there exists a set of  $(W, \mathcal{N}_n)$  with positive measure such that the true conditional outcome distribution  $\mathbf{P}(G = g | W, \mathcal{N}_n)$  puts positive probability on a set of  $\hat{g} \notin \mathcal{G}_{\theta_0, r_n}(W, \mathcal{N}_n)$  such that for some  $i, j \in \mathcal{N}_n$ , it is the case that  $\Pi_{ij}(r_n) = 1$  and either

1.  $S_{ij}(\hat{g}, W)$  is not stable with respect to  $W_{ij}$  and  $\theta_0$ , but  $\hat{g}_{ij} = 1$ , or

2.  $S_{ij}(\hat{g}, W)$  is stable with respect to  $W_{ij}$  and  $\theta_0$ , but  $\hat{g}_{ij} = 0$ .

Hence, the dyadic SM induced by the true conditional outcome distribution,

$$\mathbf{E} \left[ \sum_{g \in \mathbf{G}_{ij}(y; W)} \mathbf{P}(G = g \mid W, \mathcal{N}_n) \Big| W_{ij} \right], \quad (41)$$

must put positive probability on some  $y \notin \mathcal{S}_{\theta_0}(W_{ij}, r_n)$ . Since  $\mathbf{P}(Y_{ij} = y \mid W_{ij}) = (41)$ , this contradicts the requirement that any dyadic SM rationalizing  $\mathbf{P}(Y_{ij} = y \mid X_{ij})$  has support  $\mathcal{S}_{\theta_0}(W_{ij}, r_n)$ . ■

## References

- Andrews, D. and X. Shi**, “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 2013, *81* (2), 609–666. [34](#)
- Barabási, A.**, *Network Science*, Cambridge University Press, 2015. [2](#), [25](#)
- Barabási, A. and R. Albert**, “Emergence of Scaling in Random Networks,” *Science*, 1999, *286* (5439), 509–512. [7](#)
- Beresteanu, A., I. Molchanov, and F. Molinari**, “Sharp Identification Regions in Models with Convex Moment Inequalities,” *Econometrica*, 2011, *79* (6), 1785–1821. [30](#), [32](#), [34](#), [49](#)
- Bickel, P., A. Chen, and E. Levina**, “The Method of Moments and Degree Distributions for Network Models,” *The Annals of Statistics*, October 2011, *39* (5), 2280–2301. [12](#)
- and —, “A Nonparametric View of Network Models and Newman–Girvan and Other Modularities,” *Proceedings of the National Academy of Sciences*, 2009, *106* (50), 21068–21073. [12](#)
- Bollobás, B., S. Janson, and O. Riordan**, “The Phase Transition in Inhomogeneous Random Graphs,” *Random Structures and Algorithms*, 2007, *31* (1), 3–122. [12](#), [27](#), [48](#)
- Boucher, V. and I. Mourifié**, “My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks,” *working paper*, 2013. [4](#), [6](#)
- 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. [2](#)
- Carrell, S., B. Sacerdote, and J. West**, “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation,” *Econometrica*, 2013, *81* (3), 855–882. [2](#)

- Chandrasekhar, A. and M. Jackson**, “A Network Formation Model Based on Subgraphs,” *working paper*, 2015. [4](#)
- Chernozhukov, V., H. Hong, and E. Tamer**, “Estimation and confidence regions for parameter sets in econometric models1,” *Econometrica*, 2007, *75* (5), 1243–1284. [34](#), [37](#)
- Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman**, “An Empirical Model for Strategic Network Formation,” *working paper*, 2010. [4](#), [6](#), [7](#)
- de Paula, A.**, “Econometric Analysis of Games with Multiple Equilibria,” *Annu. Rev. Econ.*, 2013, *5* (1), 107–131. [2](#)
- , **S. Richards-Shubik, and E. Tamer**, “Identification of Preferences in Network Formation Games,” *working paper*, 2015. [4](#), [6](#), [21](#), [30](#)
- Dzemski, A.**, “An Empirical Model fo Dyadic Link Formation in a Network with Unobserved Heterogeneity,” *working paper*, 2014. [4](#)
- Fafchamps, M. and F. Gubert**, “The Formation of Risk-Sharing Networks,” *Journal of Economic Development*, 2007, *83*, 326–350. [2](#)
- Galichon, A. and M. Henry**, “Set Identification in Models with Multiple Equilibria,” *Review of Economic Studies*, 2011, *78* (4), 1264–1298. [30](#)
- Goldsmith-Pinkham, P. and G. Imbens**, “Social networks and the identification of peer effects,” *Journal of Business & Economic Statistics*, 2013, *31* (3), 253–264. [6](#), [7](#)
- Graham, B.**, “An Empirical Model of Network Formation: Detecting Homophily when Agents Are Heterogeneous,” *working paper*, 2014. [4](#), [7](#)
- Hoff, P., A. Raftery, and M. Handcock**, “Latent Space Approaches to Social Network Analysis,” *Journal of the American Statistical Associatio*, 2002, *97* (460), 1090–1098. [11](#)
- and **M. Ward**, “Modeling Dependencies in International Relations Networks,” *Political Analysis*, 2004, *12* (2), 160–175. [11](#)
- 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. [4](#)
- Jackson, M.**, *Social and Economic Networks*, Princeton University Press, 2010. [7](#), [25](#)
- and **A. Watts**, “The Evolution of Social and Economic Networks,” *Journal of Economic Theory*, October 2002, *106* (2), 265–295. [12](#), [14](#)
- Kingman, J.**, *Poisson processes*, Vol. 3, Oxford University Press, 1992. [22](#), [33](#), [42](#), [43](#)
- Leung, M.**, “Two-Step Estimation of Network-Formation Models with Incomplete Information,” *Journal of Econometrics*, 2015, *188* (1), 182–195. [4](#)

- Meester, R. and R. Roy**, *Continuum Percolation* number 119, Cambridge University Press, 1996. [11](#), [28](#)
- Mele, A.**, “A Structural Model of Segregation in Social Networks,” *working paper*, 2015. [4](#), [6](#)
- Miyauchi, Y.**, “Structural Estimation of a Pairwise Stable Network Formation with Non-negative Externality,” *working paper*, 2013. [4](#)
- Penrose, M.**, *Random Geometric Graphs*, Oxford University Press, 2003. [10](#), [26](#), [27](#)
- Penrose, Mathew D. and J. E. Yukich**, “Weak Laws of Large Numbers in Geometric Probability,” *Annals of Applied Probability*, 2003, pp. 277–303. [3](#), [24](#), [39](#), [42](#)
- and **Joseph E. Yukich**, “Central Limit Theorems for Some Graphs in Computational Geometry,” *Annals of Applied probability*, 2001, pp. 1005–1041. [42](#)
- Sheng, S.**, “Identification and Estimation of Network Formation Games,” *working paper*, 2014. [4](#), [6](#), [12](#), [13](#), [30](#)
- Sorenson, O. and T. Stuart**, “Syndication Networks and the Spatial Distribution of Venture Capital Investments,” *American journal of sociology*, 2001, *106* (6), 1546–1588. [11](#)
- Treier, S. and S. Jackman**, “Democracy as a Latent Variable,” *American Journal of Political Science*, 2008, *52* (1), 201–217. [11](#)
- Uetake, K.**, “Strategic Network Formation and Performance in the Venture Capital Industry,” *working paper*, 2012. [11](#)
- Wan, Y.**, “An Integration-Based Approach to Moment Inequality Models,” *working paper*, 2013. [34](#)
- Yildiz, Nese**, “Consistency of Plug-In Estimators of Upper Contour and Level Sets,” *Economic Theory*, 2012, *28* (02), 309–327. [34](#)