Normal Approximation in Large Network Models*

Michael P. Leung[†] Hyungsik Roger Moon[‡]

March 10, 2026

Abstract. We prove a central limit theorem for network formation models with strategic interactions and homophilous agents. Since data often consists of observations on a single large network, we consider an asymptotic framework in which the network size diverges. We argue that a modification of "stabilization" conditions from the literature on geometric graphs provides a useful high-level formulation of weak dependence which we utilize to establish an abstract central limit theorem. Using results in branching process theory, we derive interpretable primitive conditions for stabilization. The main conditions restrict the strength of strategic interactions and equilibrium selection mechanism. We discuss practical inference procedures justified by our results.

JEL Codes: C31, C57, D85

Keywords: social networks, strategic interactions, weak dependence, network formation

*We thank the editor and referees for valuable comments and suggestions that helped improve the exposition of the paper. We gratefully acknowledge financial support from the National Science Foundation under grants SES-1755100 and SES-1625586.

[†]Department of Economics, UC Santa Cruz. E-mail: leungm@ucsc.edu.

[‡]Department of Economics, University of Southern California and Yonsei University. E-mail: moonr@usc.edu.

1 Introduction

Network models have attracted considerable attention as tractable representations of non-market interactions, such as peer effects and social learning, and of economic relationships, such as financial and trade networks. The economic perspective on networks emphasizes the importance of strategic interactions or externalities (Jackson et al., 2021). One strand of the literature studies social interactions, how an individual's behavior interacts with those of her social contacts. A second strand studies network formation, why an individual chooses particular social contacts, and how their choices affect those of others. In both cases, externalities generate a wedge between choices that are optimal from the individual's perspective and those that are efficient for society. This has important consequences for policymaking, for instance motivating policies of "associational redistribution" that intervene on the network structure (Durlauf, 1996). However, when preferences are misaligned with the policy objective, the endogenous response to a policy intervention may diverge from the intended outcome (Carrell et al., 2013). It is therefore of interest to develop econometric methodologies for measuring network externalities.

The focus of our paper is models of network formation, which have diverse applications ranging from risk sharing in the rural Philippines (Fafchamps and Gubert, 2007) to research partnerships in the biotechnology industry (Powell et al., 2005). They can be used to distinguish between different theoretical mechanisms for link formation, including preferential attachment (Barabási and Albert, 1999), strategic transitivity (Mele, 2022; Ridder and Sheng, 2022), and homophily (McPherson et al., 2001). They enable policymakers to forecast the effects of counterfactual interventions on a social network (Mele, 2020) and to account for dependence between the network and unobservables of a social interactions model (Badev, 2021).

We augment a class of latent space (Hoff et al., 2002) and geometric graph (Penrose and Yukich, 2003) models with strategic interactions. In such models, nodes are positioned in a possibly latent social or economic space such that the rate of link formation decays with distance in space. However, due to strategic interactions, link formation depends not only on the attributes of the node pair involved in the link but also on links between other pairs. This induces a non-standard form of cross-sectional dependence between network units. The challenge for large-sample theory lies in establishing conditions under which the "amount of independent information" increases with the network size.

We establish a CLT for a large class of network moments computed from a single large network. The moments are averages of node-level statistics $n^{-1}\sum_{i=1}^n \psi_i$, where ψ_i is some function of the network and node types. Examples include the *degree* of node i, which is the number of links involving i; the average clustering coefficient; and subnetwork counts which have been utilized for inference in econometric models of network formation (Sheng, 2020). We discuss practical inference procedures justified by the result.

We emphasize two technical contributions. The first is an abstract CLT that holds under high-level conditions, which is an extension of limit theorems from the literature on geometric graphs (e.g. Penrose, 2007). The key condition is a modification of "stabilization" assumptions from that literature that provide a useful formulation of weak dependence for our purposes. Stabilization essentially requires ψ_i to only be a function of a random subset of nodes whose size has a distribution with exponential tails. This implies that the size is asymptotically bounded, so node i's statistic only depends on a small fraction of alters. In this sense, $\{\psi_i\}_{i=1}^n$ is weakly dependent.

In our applications, this random subset involves a union of network components, which are disconnected subnetworks that are challenging combinatorial objects to analyze (a formal definition is provided at the end of this section). Our primary contribution is to develop a methodology for deriving the tail bounds required for stabilization. We employ techniques from branching process theory commonly used to bound component sizes in the literature on random graphs. Using the methodology, we derive interpretable primitive sufficient conditions for stabilization in models of strategic network formation.

One of the main conditions restricts the strength of strategic interactions. These induce cross-sectional dependence since the realization of a link can depend on the existence of "neighboring" links, which in turn can depend on other links, and so on. The longer the lengths of these chains of dependent links, the stronger the degree of cross-sectional dependence. We adopt a well-known technique in random graph theory to bound the lengths of these chains by branching processes whose sizes can more tractably be proven to have exponential tails, provided the processes are "subcritical" or non-explosive in growth. We argue that conditions for subcriticality in our context are analogous to weak dependence conditions imposed on spatial and temporal autoregressive models that bound the magnitude of the autoregressive parameter below one.

The other main condition is a restriction on the equilibrium selection mechanism. Even if strategic interactions are sufficiently weak, strong cross-sectional dependence may exist if nodes can globally "coordinate" on the equilibrium network through a common signal such as the type of a particular node. We instead require equilibrium selection to be sufficiently "decentralized," which holds for instance under myopic best-response dynamics. This is the single-network analog of the usual requirement under many-network asymptotics that equilibrium selection is independent across networks.

A growing literature studies frequentist inference in network formation models when the econometrician observes a single network. Leung (2019b) and Menzel (2024) develop laws of large numbers for models of strategic network formation. The former paper modifies a weaker stabilization condition due to Penrose and Yukich (2003) and uses branching processes to derive primitive sufficient conditions. Our paper tackles the more difficult problem of obtaining a normal approximation, which naturally requires a stronger stabilization condition, and discusses practical inference procedures newly justified by the result. We prove the CLT by adapting results from the literature on geometric graphs (Penrose and Yukich, 2008; Penrose, 2007); a more detailed discussion of our contributions relative to this literature can be found in §3.3. Kuersteiner (2019) takes a different approach, using a novel conditional mixingale type assumption defined in terms of a random metric of distance.

Leung (2015) and Ridder and Sheng (2022) study strategic network formation under incomplete information. In this setting, links are independent conditional on observables, whereas the models we study can be microfounded as games of complete information. These allow for unobserved heterogeneity, which generates dependence between potential links even conditional on observables.

Charbonneau (2017), Dzemski (2019), Graham (2017), and Jochmans (2018) consider dyadic link formation models without strategic interactions but allow for node-level fixed effects. A large literature in statistics studies models without strategic interactions, for example stochastic block models (Bickel et al., 2011) and latent-space models (Hoff et al., 2002). These are useful for their parsimony and tasks such as community detection. Chandrasekhar and Jackson (2021) and Boucher and Mourifié (2017) study statistical models allowing for interdependence between links.

Kojevnikov et al. (2021) prove a CLT for node-level data conditional on the network. This does not apply to network formation models since the network is the outcome. We prove an unconditional CLT that may be applied to network formation as well as network processes (Leung, 2019a).

The next section presents a model of strategic network formation and defines network moments. In §3, we state high-level conditions for a CLT and outline its proof. Readers interested in the low-level CLT for strategic network formation may skip §3 for §4, which presents primitive sufficient conditions. We also outline a general methodology for verifying the high-level condition that can be applied to other network models. In §5, we discuss practical inference procedures. We present results from a simulation study in §6, and §7 concludes. Proofs can be found in the supplemental appendix.

We introduce standard notation and terminology for networks. We represent a network on a set of n nodes by an $n \times n$ adjacency matrix, where the ijth entry A_{ij}, termed the *potential link*, is an indicator for whether nodes i, j are connected. We assume $A_{ii} = 0$ for all nodes i, meaning that there are no self-links, and we focus on undirected networks, so $A_{ij} = A_{ji}$. For two networks $\boldsymbol{A}, \boldsymbol{A}'$, we say that $\boldsymbol{A}$ is a *subnetwork* of $\boldsymbol{A}'$ if every link in $\boldsymbol{A}$ is a link in $\boldsymbol{A}'$. A *path* in a network from node i to j is a sequence of distinct nodes starting with i and ending with j such that for each consecutive node pair k, k' in this sequence, $A_{kk'} = 1$. Its *length* is the number of links it involves. The *path distance* between two nodes $i \neq j$ in $\boldsymbol{A}$ is the length of the shortest path that connects them if a path exists and ∞ otherwise. The path distance between a node and itself is defined as zero. The K*-neighborhood* of a node i in $\boldsymbol{A}$, denoted by $\mathcal{N}_{\boldsymbol{A}}(i, K)$, is the set of nodes of path distance at most K from i. Finally the *component* of a node i in a network $\boldsymbol{A}$ is the set of nodes at finite path distance from i.

2 Model

Let $\mathcal{N}_n = \{1, \dots, n\}$ be a set of nodes, and endow each $i \in \mathcal{N}_n$ with an i.i.d. vector-valued *type* $(X_i, Z_i) \in \mathbb{R}^d \times \mathbb{R}^{d_z}$. We distinguish X_i as the *position* of node i, a continuously distributed vector of *homophilous* (defined below) attributes with density f that has bounded support. Endow each node pair (i, j) with an i.i.d. $\mathbb{R}$-valued random utility shock $\zeta_{ij} = \zeta_{ji}$ independent of types.

The network $\boldsymbol{A}$ satisfies, for all $i \neq j$,

$$A_{ij} = \mathbf{1}\{V_{ij} > 0\} \quad \text{for} \quad V_{ij} \equiv V(r_n^{-1}\|X_i - X_j\|, S_{ij}, Z_i, Z_j, \zeta_{ij}), \tag{1}$$

where the *joint-surplus function* $V(\cdot)$ is an $\mathbb{R}$-valued function, $\|\cdot\|$ a norm on $\mathbb{R}^d$, and S_{ij} a vector of statistics that captures strategic interactions through its dependence on $\boldsymbol{A}$. As discussed below, the sparsity of $\boldsymbol{A}$ will in part be determined by the positive constant r_n. Model (1) corresponds to the well-known *pairwise-stability* solution concept under transferable utility (Jackson, 2010).

Example 1. Consider the linear joint surplus function

$$V_{ij} = \theta_1 + \theta_2 S_{ij} - \theta_3 (r_n^{-1}\|X_i - X_j\|)^2 + \zeta_{ij} \quad \text{with} \quad S_{ij} = \max_k A_{ik}A_{jk}. \tag{2}$$

Our theory will require $\theta_3 > 0$ which captures homophily in position since it disincentivizes link formation between positionally dissimilar nodes. The term S_{ij} is an indicator for whether i and j share a common neighbor (used for example by Menzel, 2024). If $\theta_2 > 0$, then node pairs (i,j) sharing a common neighbor are more likely to form a link. Both rationalize the well-known stylized fact that networks are commonly *clustered* in that nodes with common neighbors are themselves typically neighbors (Jackson, 2010).

Example 2. Sheng (2020) studies a specification similar to

$$V_{ij} = \beta_0 + (Z_i + Z_j)'\beta_1 + \beta_2 r_n^{-1}\|X_i - X_j\| + \gamma_1 \sum_{k=1}^n (A_{ik} + A_{jk}) + \gamma_2 \sum_{k=1}^n A_{ik}A_{jk} + \zeta_{ij}, \tag{3}$$

which corresponds to linear joint surplus function with

$$S_{ij} = \left(\sum_{k=1}^n A_{ik}A_{jk}, \sum_{k=1}^n A_{ik}, \sum_{k=1}^n A_{jk} \right). \tag{4}$$

The first component of (4) plays a role analogous to S_{ij} in (2), being a count of the number of common neighbors. The second and third components are respectively the *degrees* (number of neighbors) of i and j. If $\gamma_1 > 0$, this captures a popularity effect.

As explained in the next subsection, the assumptions we impose for a CLT in many cases require S_{ij} to be uniformly bounded over $i, j \in \mathcal{N}_n$ and $n \in \mathbb{N}$. Existing work on large-network asymptotics for network formation shares this limitation (Boucher and Mourifié, 2017; Menzel, 2024; Ridder and Sheng, 2022). Statistics such as (4) may be modified to satisfy uniform boundedness by truncation, for example $\min\{\sum_{k=1}^n A_{ik}, \Delta\}$ for some user-specified constant Δ.

The next subsection states a restriction on $V(\cdot)$ that generates homophily and sparsity. In §2.2, we formalize how S_{ij} may depend on $\boldsymbol{A}$ and introduce the equilibrium selection mechanism. Finally, in §2.3, we define the class of network moments for which we provide a CLT.

2.1 Homophily and Sparsity

A common feature of social networks is *homophily*, the tendency for those with similar characteristics to associate. We require homophily in position, specifically that the joint-surplus function $V(\cdot)$ is decreasing in the first component, so nodes dissimilar in position are less likely to form links. Positions may represent latent node locations in an abstract "social space," as in latent-space models (Hoff et al., 2002), or attributes such as income and geographic location.

Another common feature is *sparsity*, meaning that the number of connections formed by the typical node is of significantly smaller order than n (Barabási, 2015). This is often accomplished by scaling the sequence of models such that the expected degree $n^{-1}\sum_{i=1}^n \sum_{j=1}^n \mathbf{E}[A_{ij}]$ is asymptotically bounded.

Our first assumption restricts the joint-surplus function $V(\cdot)$ and the distribution of ζ_{ij}, so that the network generated by model (1) exhibits homophily in positions and sparsity. Let $\bar{\Phi}_\zeta(z) = \mathbf{P}(\zeta_{ij} \geqslant z)$. Define $\bar{V}(r_n^{-1}\|X_i - X_j\|, \zeta_{ij}) = \sup_{s,z,z'} V(r_n^{-1}\|X_i - X_j\|, s, z, z', \zeta_{ij})$, where the supremum is taken over (s, z, z') in the support of (S_{ij}, Z_i, Z_j) (which will be identical across i, j under assumptions below). Finally, recall that d is the dimension of X_1.

Assumption 1 (Homophily and Sparsity)**.** *For any $\delta \in \mathbb{R}_+$, $\bar{V}(\delta, \cdot)$ is invertible and increasing in its second argument, and its inverse $\bar{V}^{-1}(\delta, \cdot)$ satisfies $\limsup_{\delta\to\infty} \delta^{-1} \log \bar{\Phi}_\zeta(\bar{V}^{-1}(\delta, 0)) < 0$. Furthermore, there exists $\kappa > 0$ such that, for any $n \in \mathbb{N}$,*

$$r_n \equiv (\kappa/n)^{1/d}. \tag{5}$$

Homophily is a consequence of the first sentence because $\mathbf{P}(A_{ij} = 1 \mid r_n^{-1}\|X_i - X_j\| = \delta) \leqslant \bar{\Phi}_\zeta(\bar{V}^{-1}(\delta, 0))$, which is required to decrease exponentially with δ. If $V(\cdot)$ is not too nonlinear in its first component and the distribution of ζ_{ij} has exponential tails, then this assumption is satisfied, as shown in the next example.

Example 3. Consider Example 1, and suppose S_{ij} has uniformly bounded support. Then for some universal constant M,

$$\mathbf{P}(V_{ij} > 0 \mid r_n^{-1}\|X_i - X_j\| = \delta) \leqslant \mathbf{P}(\underbrace{M - \theta_3\delta^2 + \zeta_{ij}}_{\bar{V}(\delta,\zeta_{ij})} > 0) = \bar{\Phi}_\zeta(\underbrace{\theta_3\delta^2 - M}_{\bar{V}^{-1}(\delta,0)}),$$

which decays to zero exponentially with δ if $\theta_3 > 0$ and the distribution of ζ_{ij} has exponential tails.

Example 4. In the previous example, Assumption 1 implicitly imposes restrictions on the support of $(S_{ij}, Z_i, Z_j, \zeta_{ij})$. Consider a variant of (2) in which we replace $\theta_3(r_n^{-1}\|X_i - X_j\|)^2$ with the "random geometric graph" penalty $\infty \cdot \mathbf{1}\{r_n^{-1}\|X_i - X_j\| > c\}$ for some $c > 0$, with the convention that $\infty \cdot 0 = 0$. That is, nodes do not link with alters whose scaled positions are sufficiently far from the ego's. Then $\mathbf{P}(V_{ij} > 0 \mid r_n^{-1}\|X_i - X_j\| = \delta) \leqslant \mathbf{1}\{\delta \leqslant c\}$ which satisfies Assumption 1 without support restrictions.

Lastly we discuss the role of (5) for sparsity. As n increases, there are more opportunities to form links, which increases expected degree, corresponding to a denser network. By sending r_n to zero with n, we increase the "cost" of link formation due to homophily, thereby decreasing the expected degree. Our choice of r_n balances these two forces to achieve a sparse network in which the expected degree is asymptotically bounded. To see this, notice that expected degree is

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbf{E}[A_{ij}] &= (n-1)\mathbf{P}\big(V(r_n^{-1}\|X_i - X_j\|, S_{ij}, Z_i, Z_j, \zeta_{ij}) > 0\big) \\
&\leqslant (n-1)\mathbf{P}\big(\zeta_{ij} > \bar{V}^{-1}(r_n^{-1}\|X_i - X_j\|, 0)\big) \\
&= (n-1)r_n^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{P}\big(\zeta_{ij} > \bar{V}^{-1}(\|x - x'\|, 0)\big) f(x) f(x + r_n(x' - x))\,\mathrm{d}x\,\mathrm{d}x' \quad (6)
\end{aligned}$$

by a change of variables $x' \mapsto x + r_n(x' - x)$. By (5), if f is continuous, this converges to

$$\kappa \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \bar{\Phi}_\zeta\big(\bar{V}^{-1}(\|x-x'\|, 0)\big) f(x)^2 \, \mathrm{d}x' \, \mathrm{d}x,$$

which is finite because the integrand decays exponentially with $\|x-x'\|$ by Assumption 1.

2.2 Strategic Interactions and Equilibrium Selection

We next define the statistics S_{ij} that capture strategic interactions. For any $H \subseteq \mathcal{N}_n$, let $r_n^{-1}\boldsymbol{T}_H = ((r_n^{-1}X_i, Z_i))_{i\in H}$ be the array of (scaled) types for nodes in H and $\boldsymbol{\zeta}_H = (\zeta_{ij})_{i,j\in H}$ (with $\zeta_{ii} \equiv 0$ for all i) be the corresponding array of random-utility shocks. In the case where $H = \mathcal{N}_n$, we abbreviate $r_n^{-1}\boldsymbol{T}_n \equiv r_n^{-1}\boldsymbol{T}_{\mathcal{N}_n}$ and $\boldsymbol{\zeta}_n \equiv \boldsymbol{\zeta}_{\mathcal{N}_n}$. Let $\boldsymbol{A}_H = (A_{ij})_{i,j\in H}$ denote the subnetwork of $\boldsymbol{A}$ on H. We define

$$S_{ij} = S_n(i, j, r_n^{-1}\boldsymbol{T}_n, \boldsymbol{A}) \tag{7}$$

where $\{S_n\}_{n\in\mathbb{N}}$ is a sequence of permutation-invariant functions.[1]

We next impose the common restriction that S_{ij} only depends on its arguments through the 1-neighborhoods of i and j, recalling the definition from §1.

Assumption 2 (Local Externalities)**.** *For any $r > 0$, $n \in \mathbb{N}$, $i, j \in \mathcal{N}_n$, and $H = \mathcal{N}_{\boldsymbol{A}}(i,1) \cup \mathcal{N}_{\boldsymbol{A}}(j,1)$, $S_n(i,j,r^{-1}\boldsymbol{T}_n, \boldsymbol{A}) = S_{|H|}(i,j,r^{-1}\boldsymbol{T}_H, \boldsymbol{A}_H)$.*

That is, S_{ij} is only a function of nodes linked to either i or j. Most statistics used in the literature satisfy this assumption, including those in Examples 1 and 2, which appear to be the most common choices (Christakis et al., 2020; Leung, 2019b; Mele, 2017; Sheng, 2020). Dependence of $S_n(\cdot)$ on types also allows for covariate-weighted versions of these examples, such as the weighted degree $\sum_{k=1}^n A_{ik}Z_k$.

The model thus far is incomplete since multiple networks $\boldsymbol{A}$ may satisfy (1) due to strategic interactions. Let $\mathcal{E}(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$ denote the set of such networks, which corresponds to the set of pairwise stable equilibria. To complete the model, we introduce a *selection mechanism*, which is the reduced-form mapping from primitives

[1] For any bijection $\pi\colon \mathcal{N}_n \to \mathcal{N}_n$, we abuse notation by defining $\pi(r_n^{-1}\boldsymbol{T}_n) = ((r_n^{-1}X_{\pi(i)}, Z_{\pi(i)}))_{i\in\mathcal{N}_n}$ and $\pi(\boldsymbol{A}) = (A_{\pi(i)\pi(j)})_{i,j}$. We say $S_n(\cdot)$ is *permutation-invariant* if $S_n(i,j,r_n^{-1}\boldsymbol{T}_n,\boldsymbol{A}) = S_n(\pi(i),\pi(j),\pi(r_n^{-1}\boldsymbol{T}_n),\pi(\boldsymbol{A}))$. This means that the function does not directly depend on node labels which have no economic content in our model.

to the observed network. It is a representation of the latent social process by which nodes coordinate on an equilibrium.

Assumption 3 (Equilibrium Selection). *For any $r > 0$ and $n \in \mathbb{N}$, an equilibrium exists in that $\mathcal{E}(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$ is non-empty, and there exists a permutation-equivariant function $\lambda_n(\cdot)$ (the* equilibrium selection mechanism*) such that $\boldsymbol{A} = \lambda_n(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n) \in \mathcal{E}(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$.*[2]

Remark 1. The empirical games literature typically represents the equilibrium selection mechanism as a conditional distribution $\sigma(\cdot \mid r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$ over $\mathcal{E}(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$. We represent it as a deterministic function, which is more convenient for our purposes, especially for formulating Assumption 8 below. We next show that for any conditional distribution, we can construct a deterministic function that induces the same distribution over equilibria. The main idea is that types can always include payoff-irrelevant components that can be used to randomize over equilibria. Specifically, let $\{\nu_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{U}([0,1])$ be independent of all structural primitives. These will serve to generate a distribution over equilibria conditional on structural primitives. Redefine $r^{-1}\boldsymbol{T}_n = ((r^{-1}X_i, Z_i, \nu_i))_{i\in\mathcal{N}_n}$ (scaled types with payoff-irrelevant components), and let $r^{-1}\tilde{\boldsymbol{T}}_n = ((r^{-1}X_i, Z_i))_{i\in\mathcal{N}_n}$ (scaled types as originally defined).

The set of equilibria is given by $\mathcal{E}(r^{-1}\tilde{\boldsymbol{T}}_n, \boldsymbol{\zeta}_n) = \{\boldsymbol{A}_k\}_{k=1}^m$. Consider any conditional distribution σ over this set, where we abbreviate $\sigma_k \equiv \sigma(\boldsymbol{A}_k \mid r^{-1}\tilde{\boldsymbol{T}}_n, \boldsymbol{\zeta}_n)$. Define $\nu^* = F(1 - \max\{\nu_i\}_{i=1}^n)$, where F is the CDF of $1 - \max\{\nu_i\}_{i=1}^n$. Partition $[0,1]$ into m intervals of lengths $\sigma_1, \dots, \sigma_m$, and let $\lambda_n(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$ be the function that selects equilibrium $\boldsymbol{A}_k$ if ν^* falls within the interval associated with σ_k. Since $\nu^* \sim \mathcal{U}([0,1])$ for any n, we have $\mathbf{P}(\lambda_n(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n) = \boldsymbol{A}_k \mid r^{-1}\tilde{\boldsymbol{T}}_n, \boldsymbol{\zeta}_n) = \sigma_k$. In other words, our deterministic construction of $\lambda_n(\cdot)$ generates the desired conditional distribution over equilibria.

[2] Following the notation in footnote 1, we say $\lambda_n(\cdot)$ is permutation-equivariant if $\pi(\lambda_n(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)) = \lambda_n(\pi(r^{-1}\boldsymbol{T}_n), \pi(\boldsymbol{\zeta}_n))$, where $\pi(\boldsymbol{\zeta}_n)$ is defined similarly to $\pi(\boldsymbol{A})$.

2.3 Network Moments

Our objective is to prove a CLT for network moments that are averages of *node statistics*

$$\frac{1}{n}\sum_{i=1}^{n}\psi_n(i, r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n, \boldsymbol{A}),$$

where $\{\psi_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathbb{R}^{d_\psi}$-valued, permutation-invariant functions (see footnote 1). We often abbreviate $\psi_i(\mathcal{N}_n) \equiv \psi_n(i, r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n, \boldsymbol{A})$.

Example 5 (Subnetwork Counts)**.** A simple network moment is the *average degree* for which the node statistic is the degree $\psi_i(\mathcal{N}_n) = \sum_{j=1}^n A_{ij}$. This is a permutation-invariant function of the network. Average degree is proportional to the dyad count (number of links). More generally, we can count any other connected subnetwork, such as the number of triangles, k-stars, or complete networks on k-tuples. For instance, the triangle count is proportional to $\sum_{i,j,k} A_{ij}A_{jk}A_{ik}$, with corresponding node statistic $\psi_i(\mathcal{N}_n) = \sum_{j,k} A_{ij}A_{jk}A_{ik}$. See §A for a formal definition of subnetwork counts.

More generally, we consider the following class of node statistics that includes the previous examples. Recall that $\mathcal{N}_{\boldsymbol{A}}(i,K)$ is i's K-neighborhood, defined in §1.

Assumption 4 (Node Statistics)**.** *For some $K \in \mathbb{Z}_+$ and any $r > 0$, $n \in \mathbb{N}$, $i \in \mathcal{N}_n$, and $H = \mathcal{N}_{\boldsymbol{A}}(i,K)$, $\psi_n(i, r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n, \boldsymbol{A}) = \psi_{|H|}(i, r^{-1}\boldsymbol{T}_H, \boldsymbol{\zeta}_H, \boldsymbol{A}_H)$.*

This states that the node statistic only depends on the types, random-utility shocks, and subnetwork of nodes on i's K-neighborhood. The average degree and triangle count satisfy this for $K = 1$. A more complex example is the number of nodes at most path distance D from node i, which satisfies this assumption for $K = D$.

Example 6. Subnetwork counts are useful for structural inference. Suppose $V(\cdot)$ is known up to some vector of parameters θ_0. Sheng (2020) defines an identified set for θ_0 in terms of moment inequalities of the form

$$\frac{1}{n}\sum_{i_1=1}^{n}\dots\sum_{i_m=1}^{n}\mathbf{E}[(G_{i_1,\dots,i_m} - H_{i_1,\dots,i_m}(\theta_0))q_{i_1,\dots,i_m}(\boldsymbol{T}_n)] \leqslant 0.^3 \tag{8}$$

[3] Technically this is proportional to the expectation of (5.3) in Sheng (2020). Our scaling is

Here $\sum_{i_1=1}^n \dots \sum_{i_m=1}^n G_{i_1,\dots,i_m}$ is proportional to a count of a particular connected subnetwork of size m. For example, for counting triangles, which are fully connected subnetworks with $m = 3$, $G_{i_1,\dots,i_m} = A_{i_1 i_2} A_{i_2 i_3} A_{i_3 i_1}$. The upper bound $H_{i_1,\dots,i_m}(\theta_0)$ and instrument function $q_{i_1,\dots,i_m}(\boldsymbol{T}_n)$ are known, deterministic functions of the observed component of $\boldsymbol{T}_n$.

Furthermore, $n^{-1} \sum_{i_1=1}^n \dots \sum_{i_m=1}^n H_{i_1,\dots,i_m}(\theta_0) q_{i_1,\dots,i_m}(\boldsymbol{T}_n)$ is proportional to a U-statistic of order m with a kernel that is a deterministic function of $\{(X_{i_k}, Z_{i_k})\}_{k=1}^m$. By the Hoeffding decomposition, it equals $n^{-1} \sum_{i=1}^n J_i(\theta_0) + o_p(n^{-1/2})$ for some $J_i(\theta_0)$ with the same mean that is a deterministic function of i's type (X_i, Z_i). Hence, $J_i(\theta_0)$ satisfies Assumption 4 for $K = 0$. Additionally, $\tilde{\psi}_{i_1}(\mathcal{N}_n) \equiv \sum_{i_2=1}^n \dots \sum_{i_m=1}^n G_{i_1,\dots,i_m} q_{i_1,\dots,i_m}(\boldsymbol{T}_n)$ is a node statistic satisfying Assumption 4 for some $K \leqslant m - 1$ since the subnetwork is connected. We have therefore shown that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i_1=1}^n \dots \sum_{i_m=1}^n \big[&(G_{i_1,\dots,i_m} - H_{i_1,\dots,i_m}(\theta_0)) q_{i_1,\dots,i_m}(\boldsymbol{T}_n) \\ &- \mathbf{E}[(G_{i_1,\dots,i_m} - H_{i_1,\dots,i_m}(\theta_0)) q_{i_1,\dots,i_m}(\boldsymbol{T}_n)]\big] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\underbrace{\tilde{\psi}_i(\mathcal{N}_n) - J_i(\theta_0)}_{\psi_i(\mathcal{N}_n)} - \mathbf{E}[\tilde{\psi}_i(\mathcal{N}_n) - J_i(\theta_0)] \Big) + o_p(1). \end{aligned}$$

Our CLT can be applied to the first term on the last line.

3 Stabilization

This section provides high-level conditions for a CLT. Readers interested in the application to strategic network formation may wish to skip to §4 where we state primitive sufficient conditions.

Recall from Assumption 3 that for any $H \subseteq \mathcal{N}_n$, $\lambda_{|H|}(r_n^{-1}\boldsymbol{T}_H, \boldsymbol{\zeta}_H)$ is the equilibrium network under the counterfactual model in which the set of nodes is H rather than $\mathcal{N}_n$. Abusing notation, for $i \in H$, we define i's *counterfactual node statistic*

$$\psi_i(H) \equiv \psi_{|H|}\big(i, r_n^{-1}\boldsymbol{T}_H, \boldsymbol{\zeta}_H, \lambda_{|H|}(r_n^{-1}\boldsymbol{T}_H, \boldsymbol{\zeta}_H)\big). \tag{9}$$

For $H = \mathcal{N}_n$, this coincides with the original definition. For $H \subset \mathcal{N}_n$, the counterfac-

different because we consider sparse networks.

tual interpretation is due entirely to the last argument. Contrast (9) with

$$\psi_{|H|}(i, r_n^{-1}\boldsymbol{T}_H, \boldsymbol{\zeta}_H, \boldsymbol{A}_H). \tag{10}$$

For example, if $\psi_i(\mathcal{N}_n)$ is the degree of node i on the full network, then (10) is i's degree in the observed subnetwork restricted to H, which can be directly computed from observation of $\boldsymbol{A}$. Generally, this does not equal (9) because the networks $\boldsymbol{A}_H$ and $\lambda_{|H|}(r_n^{-1}\boldsymbol{T}_H, \boldsymbol{\zeta}_H)$ do not coincide due to strategic interactions, which induce nodes in H to form different links when nodes $\mathcal{N}_n \backslash H$ are absent from the model.

Recall that d is the dimension of X_1. For any $B \subseteq \mathbb{R}^d$, let $\mathcal{N}_n(B) = \{i \in \mathcal{N}_n \colon r_n^{-1}X_i \in B\}$, the set of nodes with scaled positions lying in the set B. For $x \in \mathbb{R}^d$ and $R \geqslant 0$, let $B(x, R) = \{y \in \mathbb{R}^d \colon \|x - y\| \leqslant R\}$.

Definition 1. For any $r > 0$, $n \in \mathbb{N}$, and $i \in \mathcal{N}_n$, the *radius of stabilization*

$$\boldsymbol{R}_i(n, r) \equiv \boldsymbol{R}(i, r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n, \lambda_n)$$

is the smallest integer-valued $R \geqslant 0$ such that $\psi_i(\mathcal{N}_n) = \psi_i(\mathcal{N}_n(\mathcal{B}_i))$ for all sets $\mathcal{B}_i \subseteq \mathbb{R}^d$ containing $B(r^{-1}X_i, R)$.

This is the smallest radius R such that i's node statistic has the same value under all counterfactual models that drop nodes positioned outside of i's neighborhood $B(r^{-1}X_i, R)$.

The main idea is as follows. If this radius were "small," then $\psi_i(\mathcal{N}_n)$ would primarily depend on a small fraction of nodes, specifically those positioned near i, in which case node statistics should be weakly dependent. To establish a law of large numbers (LLN), Leung (2019b) requires the radius to be $O_p(1)$, analogous to a condition due to Penrose and Yukich (2003). However, a CLT seems to require stronger conditions on the distribution of the radius, in particular the thickness of its tails (Penrose, 2007).

3.1 Abstract CLT

We first define some notation and the sequence along which we take limits. Let $\{(X_i, Z_i)\}_{i\in\mathbb{N}}$ and $\{\zeta_{ij} \colon i \in \mathcal{N}_n, j > i\}$ both be i.i.d. and independent, and assume the density f of X_1 is bounded. We consider a sequence indexed by $n \in \mathbb{N}$ such that the

nth element of the sequence is given by the tuple

$$(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n, \lambda_n, \psi_n).$$

The first three components are structural primitives that determine $\boldsymbol{A}$, while the last is the node statistic function.

For technical reasons, we need to consider sequences of models indexed by n in which the network size is not n but rather of the same asymptotic order as n. Let $\{N_n\}_{n\in\mathbb{N}}$ be a sequence of random variables independent of all model primitives such that

$$N_n \sim \text{Poisson}(n). \tag{11}$$

A de-Poissonization argument discussed below requires us to consider two different network sizes: a random size $N_n + k$, where k is a constant, and a non-random size $m + k$, where $m = m_n$ with $m_n/n \to c \in (0, \infty)$. Our original setup with n nodes corresponds to $m = n$ and $k = 0$, and no intuition is lost to the reader who only considers this case. The need to consider models with different network sizes is purely for technical reasons clarified in §3.2.

Define $\mathcal{T} = \text{supp}(X_1, Z_1)$, $\mathcal{T}^k = \times_{i=1}^k \mathcal{T}$, and $\boldsymbol{T}_k = ((X_i, Z_i))_{i=1}^k$. Our first assumption is the key high-level weak dependence condition that controls the tail behavior of the distribution of the radius of stabilization, or more formally the conditional probability $\mathbf{P}(\boldsymbol{R}_1(m + k, r_n) > w \mid \boldsymbol{T}_k = \boldsymbol{t}_k)$. Since probabilities conditional on continuous random variables are not uniquely defined, we require some additional notation to phrase the assumption in terms of a *version* of the conditional probability. Let $m, k \in \mathbb{N}$ and $\boldsymbol{t}_k \in \mathcal{T}^k$. Construct $r_n^{-1}\boldsymbol{T}^\star_{m+k}(\boldsymbol{t}_k)$ from $r_n^{-1}\boldsymbol{T}_{m+k}$ by replacing the types of the first k nodes $\boldsymbol{T}_k$ with fixed values $\boldsymbol{t}_k$. Define $\boldsymbol{R}^\star_1(m + k, r_n; \boldsymbol{t}_k) \equiv \boldsymbol{R}(1, r_n^{-1}\boldsymbol{T}^\star_{m+k}(\boldsymbol{t}_k), \boldsymbol{\zeta}_{m+k}, \lambda_{m+k})$. Then $\mathbf{P}(\boldsymbol{R}^\star_1(m + k, r_n; \boldsymbol{t}_k) > w)$ is a version of the conditional probability $\mathbf{P}(\boldsymbol{R}_1(m + k, r_n) > w \mid \boldsymbol{T}_k = \boldsymbol{t}_k)$.

Assumption 5 (Exponential Stabilization)**.** *For any* $k \in \{1, \dots, 4\}$ *and* $k' \in \{1, 2\}$, *there exist* $n_0, \epsilon > 0$ *and* $\eta \in (0,1]$ *such that*

$$\limsup_{w\to\infty} w^{-\eta} \max\{\log \tau_{b,\epsilon}(w), \log \tau_p(w)\} < 0, \quad \textit{where}$$
$$\tau_{b,\epsilon}(w) = \sup_{n>n_0} \sup_{m\in((1-\epsilon)n,(1+\epsilon)n)} \sup_{\boldsymbol{t}_k\in\mathcal{T}^k} \mathbf{P}\big(\boldsymbol{R}_1^\star(m+k, r_n; \boldsymbol{t}_k) > w\big),$$
$$\tau_p(w) = \sup_{n>n_0} \sup_{\boldsymbol{t}_{k'}\in\mathcal{T}^{k'}} \mathbf{P}\big(\boldsymbol{R}_1^\star(N_n+k', r_n; \boldsymbol{t}_{k'}) > w\big).$$

For $\eta = 1$, this says that the distribution of the radius of stabilization has an exponential tail under models with $N_n + k$ and $m + k$ nodes, uniformly over the types of the first k nodes.[4] Technically the assumption allows for slower than exponential decay with $\eta < 1$, but it is otherwise analogous to conditions used by Penrose and Yukich (2008) and Penrose (2007). Our method of proof for the result below is based on theirs but with important differences in setup discussed in §3.3.

Contrast Assumption 5 with the more familiar concept of m-dependence, which states that an observation is only correlated with nodes in some non-random neighborhood of known radius m. Stabilization generalizes this to allow neighborhoods to be node-specific, random, and complex functions of the primitives. In general, bounding the size of this set is far from trivial, and one of our main contributions is to demonstrate that branching processes can be used for this purpose to derive primitive sufficient conditions, as discussed in §4.5.

The next assumption imposes a moment condition. Similar to the previous assumption, we require additional notation to phrase the condition in terms of a version of a conditional expectation. Let $k \in \mathbb{N}$, $\boldsymbol{t}_k \in \mathcal{T}^k$, and $H \subseteq \mathbb{N}$ be a finite set containing $\{1, \dots, k\}$. Construct $r_n^{-1}\boldsymbol{T}_H^\star(\boldsymbol{t}_k)$ from $r_n^{-1}\boldsymbol{T}_H$ by replacing the types of the first k nodes $\boldsymbol{T}_k$ with fixed values $\boldsymbol{t}_k$. Define $\psi_1^\star(H; \boldsymbol{t}_k) \equiv \psi_{|H|}(1, r_n^{-1}\boldsymbol{T}_H^\star(\boldsymbol{t}_k), \boldsymbol{\zeta}_H, \lambda_{|H|})$, so that $\mathbf{E}[\psi_1^\star(H; \boldsymbol{t}_k)]$ is a version of the conditional expectation $\mathbf{E}[\psi_1(H) \mid \boldsymbol{T}_k = \boldsymbol{t}_k]$. Finally, let $\|x\|_\infty$ denote the entry-wise maximum of a vector or matrix x.

Assumption 6. *(a) There exist* $p > 2, \epsilon > 0$, $M < \infty$, *and* $n_0 \in \mathbb{N}$ *such that for all* $n > n_0$, $k \in \{1, 2, 3\}$, $m \in ((1-\epsilon)n, (1+\epsilon)n)$, *and* $\boldsymbol{t}_k \in \mathcal{T}^k$,

$$\max\big\{\mathbf{E}[\|\psi_1^\star(\mathcal{N}_{m+k}; \boldsymbol{t}_k)\|_\infty^p], \mathbf{E}[\|\psi_1^\star(\mathcal{N}_{N_n+k} \cap (H_n \cup \{1\}); \boldsymbol{t}_k)\|_\infty^p]\big\} < M$$

[4] The first k is an arbitrary choice of k nodes. Note that nodes are exchangeable in our model since types and random-utility shocks are i.i.d. and the equilibrium selection mechanism is permutation-equivariant.

for any sequence of sets $H_n \subseteq \mathbb{N}$. (b) For any $r > 0$, $n \in \mathbb{N}$, $i \in \mathcal{N}_n$, and $x \in \mathbb{R}^d$,

$$\psi_n(i, r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n, \boldsymbol{A}) = \psi_n\big(i, ((r^{-1}X_j + x, Z_j))_{j=1}^n, \boldsymbol{\zeta}_n, \boldsymbol{A}\big) \quad \textit{and}$$
$$\lambda_n(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n) = \lambda_n\big((r^{-1}X_j + x, Z_j))_{j=1}^n, \boldsymbol{\zeta}_n\big).$$

Part (a) requires node statistics to have bounded $p > 2$ moments, uniformly over the types of k nodes. We provide primitive conditions in §A for the case of subnetwork counts (Example 5).

Part (b) says that node statistics and the selection mechanism are invariant to additive shifts in scaled positions $\{r^{-1}X_i\}_{i=1}^n$, which holds if these only enter through scaled distances $r^{-1}\|X_i - X_j\|$. Because X_i and Z_i may be arbitrarily dependent, this still allows X_i (unscaled) to enter the model as a subvector of Z_i. In Examples 1 and 2, the latent index only depends on types through scaled distances or Z_i, so requiring the selection mechanism and node statistics to satisfy the same property does not apparently rule out any economically interesting applications.[5]

Let $\boldsymbol{\Sigma}_n = n^{-1}\text{Var}(\sum_{i=1}^n \psi_i(\mathcal{N}_n))$, $\lambda_{\min}(\boldsymbol{\Sigma}_n)$ be its smallest eigenvalue, and $\boldsymbol{I}$ denote the d_ψ-dimensional identity matrix.

Theorem 1. *Under Assumptions 5 and 6,* $\sup_n \|\boldsymbol{\Sigma}_n\|_\infty < \infty$. *Further suppose that* $\liminf_{n\to\infty} \lambda_{min}(\boldsymbol{\Sigma}_n) > 0$. *Then*

$$\boldsymbol{\Sigma}_n^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n \big(\psi_i(\mathcal{N}_n) - \mathbf{E}[\psi_i(\mathcal{N}_n)]\big) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}). \tag{12}$$

Proof. See §SA.1. ∎

The proof does not rely on the setup or assumptions in §2, other than the distributional assumptions at the top of §2 and the requirement that $\lambda_n(\cdot)$ is a deterministic, permutation-equivariant function of the structural primitives. The other assumptions will be used in the next section to verify Assumption 5 for strategic network formation.[6]

[5]For example, Assumption 6(b) disallows network moments such as $n^{-1}\sum_{i=1}^n \sum_{j=1}^n r^{-1}X_j A_{ij}$, but since X_i and Z_i may be arbitrarily correlated, it allows for $n^{-1}\sum_{i=1}^n \sum_{j=1}^n X_j A_{ij}$.

[6]Theorem 1 can be applied to other network models. For instance, in Leung (2019a), $\lambda_n(\cdot)$ is the reduced-form mapping that takes as input the structural primitives and outputs both the network and the outcome of a second-stage social interactions model given the network.

3.2 Outline of Proof

Step 1. We first establish a CLT for the "Poissonized" model in which the number of nodes is N_n defined in (11), so-called because $\{X_i\}_{i=1}^{N_n}$ has the same distribution as a Poisson point process with intensity function $nf(\cdot)$ (Penrose, 2003, Proposition 1.5). Specifically, we show

$$\tilde{\boldsymbol{\Sigma}}_n^{-1/2}\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{N_n}\psi_i(\mathcal{N}_{N_n})-\mathbf{E}\left[\sum_{i=1}^{N_n}\psi_i(\mathcal{N}_{N_n})\right]\right)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{I}), \tag{13}$$

where $\tilde{\boldsymbol{\Sigma}}_n = n^{-1}\text{Var}(\sum_{i=1}^{N_n}\psi_i(\mathcal{N}_{N_n}))$. This is simpler to prove directly because, unlike $\{X_i\}_{i=1}^{n}$, the Poisson process $\{X_i\}_{i=1}^{N_n}$ possesses a well-known spatial independence property: for any disjoint subsets $S_1, S_2 \subseteq \mathbb{R}^d$, $|\{X_i\}_{i=1}^{N_n} \cap S_1| \perp\!\!\!\perp |\{X_i\}_{i=1}^{N_n} \cap S_2|$.

We prove (13) by adapting a theorem due to Penrose and Yukich (2008) (see our Theorem SA.1.1). We proceed by first partitioning the support of X_1 into cubes $Q_1, \dots, Q_{V_n}$ of slowly growing volume and representing the moment $n^{-1}\sum_{i=1}^{N_n}\psi_i(\mathcal{N}_{N_n})$ as a double sum over cubes and nodes within cubes $n^{-1}\sum_{i=1}^{V_n}\sum_{j=1}^{N_i}\psi_{ij}(\mathcal{N}_{N_n})$, where N_i is the number of nodes positioned in Q_i and ij indexes the jth node in Q_i. Spatial independence of the Poisson process implies independence of node positions across cubes. However, node statistics are complex functionals of the process and hence are not generally independent across cubes.

Since stabilization implies that node statistics $\psi_{ij}(\mathcal{N}_n)$ are primarily determined by nodes relatively proximate to ij, and hence cubes relatively proximate to Q_i, it can be shown that the dependence structure of $\{\sum_{j=1}^{N_i}\psi_{ij}(\mathcal{N}_{N_n})\}_{i=1}^{V_n}$ is "approximately" characterized by a certain "dependency graph." This is a network in which, roughly speaking, two observations are linked if and only if they are dependent. In our case, observations are cubes, which we connect if and only if they are relatively proximate. Careful construction of the cubes ensures a small approximation error for the dependency graph characterization, and applying a CLT for dependency graphs (Chen and Shao, 2004) delivers the result. For precise details, see §SA.1.1.

Step 2. Since $N_n/n \xrightarrow{p} 1$, $\{X_i\}_{i=1}^{n}$ and $\{X_i\}_{i=1}^{N_n}$ should be "similar," so given (13), we expect a similar result for the original model. The second "de-Poissonization" step of the proof, which follows Penrose (2007), shows that this intuition is correct, provided we properly adjust $\tilde{\boldsymbol{\Sigma}}_n$ downward to obtain the correct variance $\boldsymbol{\Sigma}_n$ (see our Theorem SA.1.2). This is needed because N_n contributes additional randomness to the asymptotic distribution.

More specifically, define the *add-one cost*

$$\Xi_n = \psi_{n+1}(\mathcal{N}_{n+1}) + \sum_{i=1}^{n} \big(\psi_i(\mathcal{N}_{n+1}) - \psi_i(\mathcal{N}_n)\big). \tag{14}$$

This is the aggregate counterfactual impact on the total $\sum_{i=1}^n \psi_i(\mathcal{N}_n)$ from adding a new node labeled $n+1$ to the model. The first term is the direct effect of adding $n+1$, which is its own node statistic. The second term is the indirect effect, which is the new node's impact on the statistics of all other nodes. A key step of the proof establishes that

$$\begin{aligned} n^{-1/2}\left(\sum_{i=1}^{N_n}\psi_i(\mathcal{N}_{N_n}) - \mathbf{E}\left[\sum_{i=1}^{N_n}\psi_i(\mathcal{N}_{N_n})\right]\right) \\ = n^{-1/2}\left(\sum_{i=1}^{n}\psi_i(\mathcal{N}_n) - \mathbf{E}\left[\sum_{i=1}^{n}\psi_i(\mathcal{N}_n)\right]\right) \\ + n^{-1/2}(N_n - n)\mathbf{E}[\Xi_{N_n}] + o_p(1). \end{aligned}$$

This may be viewed as a first-order expansion in the number of nodes, comparing N_n to n. The "derivative" is $\mathbf{E}[\Xi_{N_n}]$ since it captures the change in moments as a result of a unit increment in the number of nodes. By (13), the left-hand side is asymptotically normal, and by the Poisson CLT, so is $n^{-1/2}(N_n - n)\mathbf{E}[\Xi_{N_n}]$. Because N_n is independent of all other primitives, we can then establish that the first term on the right-hand side is asymptotically normal by an argument using characteristic functions; for details see the end of §SA.1.2.

3.3 Related Literature

The proof is closely based on arguments in Penrose and Yukich (2008) and Penrose (2007), whose results pertain to geometric graphs without strategic interactions. The innovation in Theorem 1 is primarily conceptual, namely, the recognition that an appropriate modification of stabilization allows us to adapt their results to econometric models. Our main technical innovation will be discussed in §4, namely the use of branching processes to derive primitive conditions for stabilization in models with strategic interactions. Leung (2019b) uses branching processes to establish an

LLN, while we tackle the more difficult task of proving a CLT, which requires us to establish new tail bounds for the radius of stabilization (Lemma SA.2.8).

The setup and assumptions used in Penrose's work are not directly applicable to our setting, so we cannot simply verify their conditions. However, we show that their proofs can be translated to our setting, which differs in three main aspects. The first is the definition of the radius of stabilization. We reformulate the definition in terms of counterfactual models (9) and require invariance of i's node statistic to the removal of nodes outside of $\mathcal{N}_n(\mathcal{B}_i)$. Existing definitions demand invariance to the removal *and addition* of new nodes, but invariance to addition is typically violated in the models we study due to strategic interactions. Second, X_i may be correlated with Z_i in our setup, whereas the literature requires independence, but this turns out to have little effect on the proofs. Third, our model includes pair-specific shocks ζ_{ij}, which pose little problem due to their high degree of independence. These are independent across pairs, in contrast to type pairs $((X_i, Z_i), (X_j, Z_j))$ which are correlated across pairs sharing a common node, for example (i, j) and (i, k).

4 CLT for Network Formation

Exponential stabilization (Assumption 5) provides a high-level formulation of weak dependence. This section derives primitive conditions for the network formation model in §2, so throughout this section we work under its setup. We begin in §4.1 by introducing key definitions used in §4.2 to explain two sources of cross-sectional dependence induced by the model. These motivate the weak dependence conditions stated in §4.3 and §4.4. In §4.5 we present the main result, that these conditions imply exponential stabilization. We outline the method of proof in §4.6.

4.1 Strategic Neighborhood

Recall the definition of the joint surplus from (1), and let

$$D_{ij} = \mathbf{1}\big\{ \sup_s V(r_n^{-1}\|X_i - X_j\|, s, Z_i, Z_j, \zeta_{ij}) > 0 \big\} \times \mathbf{1}\big\{ \inf_s V(r_n^{-1}\|X_i - X_j\|, s, Z_i, Z_j, \zeta_{ij}) \leqslant 0 \big\}. \quad (15)$$

This is an indicator for whether the potential link A_{ij} is *non-robust*. If $\inf_s V(r_n^{-1}\|X_i -$

$X_j\|, s, Z_i, Z_j, \zeta_{ij}) > 0$, then $A_{ij} = 1$, and the link is *robust* in that the joint surplus is positive regardless of what other links are formed. This is because $\boldsymbol{A}$ enters $V(\cdot)$ only through S_{ij}. Likewise, if $\sup_s V(r_n^{-1}\|X_i - X_j\|, s, Z_i, Z_j, \zeta_{ij}) \leqslant 0$, then $A_{ij} = 0$, and the link is *robustly absent* in that the joint surplus is negative regardless of what other links are formed. In either case, $D_{ij} = 0$. If instead $D_{ij} = 1$, then A_{ij} may be 0 or 1, and the potential link is non-robust in that the sign of the joint surplus is responsive to links formed by others.

Let $\boldsymbol{D}$ be the network of non-robustness indicators with ijth entry D_{ij}. Let C_i denote i's component in $\boldsymbol{D}$, recalling from §1 that a component is a connected subnetwork that is disconnected from the rest of the network. Let $\boldsymbol{\Pi}$ be the network of robust link indicators with ijth entry

$$\Pi_{ij} = \mathbf{1}\Big\{\inf_s V(r_n^{-1}\|X_i - X_j\|, s, Z_i, Z_j, \zeta_{ij}) > 0\Big\}.$$

Recall that $\mathcal{N}_{\boldsymbol{\Pi}}(i,1)$ denotes i's 1-neighborhood in $\boldsymbol{\Pi}$, which includes i itself. A crucial concept for what follows is a node's *strategic neighborhood*, given by

$$C_i^+ = \bigcup \{\mathcal{N}_{\boldsymbol{\Pi}}(j,1)\colon j \in C_i\}\,. \tag{16}$$

This adds to C_i the set of all nodes that possess a robust link to some member of C_i.

Example 7. Consider Example 1, and suppose $\theta_2 > 0$. Then A_{ij} is robust if $\theta_1 - \theta_3(r_n^{-1}\|X_i - X_j\|)^2 + \zeta_{ij} > 0$ and robustly absent if $\theta_1 + \theta_2 - \theta_3(r_n^{-1}\|X_i - X_j\|)^2 + \zeta_{ij} \leqslant 0$. The non-robust indicator is

$$D_{ij} = \mathbf{1}\Big\{-\theta_2 < \theta_1 - \theta_3(r_n^{-1}\|X_i - X_j\|)^2 + \zeta_{ij} \leqslant 0\Big\}.$$

As the strength of strategic interactions θ_2 increases, so does the right-hand side, and hence, the likelihood of non-robustness. In the case of no strategic interactions ($\theta_2 = 0$), there are no non-robust potential links.

We can compute C_i as follows. Initialize $C_i = \{i\}$, add i's neighbors in the network $\boldsymbol{D}$ to the set, and then iteratively add neighbors of neighbors in the manner of a breadth-first search until there are no new nodes to add. To compute C_i^+, we set $\Pi_{ij} = \mathbf{1}\{\theta_1 - \theta_3(r_n^{-1}\|X_i - X_j\|)^2 + \zeta_{ij} > 0\}$ for all $i \neq j$ and add to C_i all nodes that are neighbors under $\boldsymbol{\Pi}$ of some member of C_i.

The previous example illustrates how D_{ij} is increasing in the strength of strategic interactions. Stronger interactions then imply that C_i^+ is a larger set for any given realization of the primitives. This suggests that, *when strategic neighborhoods are likely small in size, we expect weaker strategic interactions and hence weaker cross-sectional dependence.* We elaborate on this point in the next subsection.

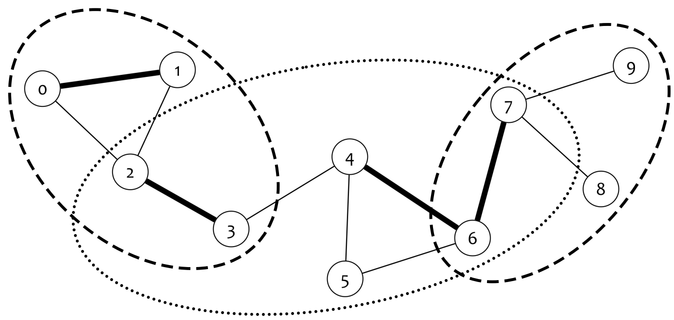

Figure 1: A network with three strategic neighborhoods.

Example 8. Let the primitives $(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$ be given, which fully determine the realizations of non-robustness indicators $\boldsymbol{D}$ and robust link indicators $\boldsymbol{\Pi}$ defined in Example 7. Suppose their realizations are depicted in Figure 1, where thin lines represent non-robust potential links (those of $\boldsymbol{D}$), solid lines represent robust links (those of $\boldsymbol{\Pi}$), and the absence of a line between two nodes represents a robustly absent link. The observed network $\boldsymbol{A}$ is a subnetwork of the depicted network because if $D_{ij} = 1$, then A_{ij} is either 1 or 0, while if $D_{ij} = 0$, then the link between i, j is either robustly absent, in which case $A_{ij} = 0$, or robust ($\Pi_{ij} = 1$), in which case $A_{ij} = 1$. The components of $\boldsymbol{D}$ are the three subnetworks obtained by removing the solid lines: $\{0, 1, 2\}$, $\{3, 4, 5, 6\}$, and $\{7, 8, 9\}$. The strategic neighborhoods are the three circled subnetworks $\{0, 1, 2, 3\}$, $\{2, \dots, 7\}$, and $\{6, 7, 8, 9\}$, obtained by adding to each component the set of nodes with solid lines to any member of the component. Notice that strategic neighborhoods may have nodes in common, whereas components necessarily partition $\mathcal{N}_n$.

4.2 Dependence Structure

Models of strategic network formation induce two forms of cross-sectional dependence, one due to link interdependencies directly induced by strategic interactions, and the other due to equilibrium selection. To illustrate, consider a model with $n = 9$ nodes and realized structural primitives such that Figure 1 depicts the resulting robust and non-robust potential links. Consider the impact on the realized network $\boldsymbol{A}$ of a hypothetical intervention that perturbs the value of ζ_{56}.

4.2.1 Best-Response Chains

The perturbation changes the joint surplus V_{56} to some new value $V_{56}^{(1)}$, both evaluated under $\boldsymbol{A}$. If the signs of these values differ, then A_{56} is no longer pairwise stable, so suppose in response the pair updates their potential link A_{56} to $A_{56}^{(1)} = \mathbf{1}\{V_{56}^{(1)} > 0\}$, resulting in the network $\boldsymbol{A}_1$. This update may affect the joint surplus of other potential links, for instance V_{67} via S_{67}. However, because A_{67} is a robust link, V_{67} is always positive, whether evaluated under $\boldsymbol{A}$ or $\boldsymbol{A}_1$, so there is no subsequent update to A_{67}. On the other hand, A_{45} is non-robust ($D_{45} = 1$), so the update to A_{56} may change the value of V_{45} evaluated under $\boldsymbol{A}$ to a new value $V_{45}^{(2)}$ evaluated under $\boldsymbol{A}_1$ that has a different sign. Suppose in response the pair updates A_{45} to $A_{45}^{(2)} = \mathbf{1}\{V_{45}^{(2)} > 0\}$, resulting in a new network $\boldsymbol{A}_2$. This in turn affects the joint surplus of any other node pair containing nodes 4 and 5. The only such pairs with non-robust potential links are $(3,4)$ and $(5,6)$ since $D_{34} = D_{56} = 1$, so *only* those potential links may update in response to the update to A_{45}.

Suppose we iterate these best-response dynamics indefinitely. At no point do potential links other than those of pairs $(3,4)$, $(4,5)$, and $(5,6)$ update during the process. This is because they are the only pairs connected to nodes 5 or 6, the target of the initial perturbation, through a chain of non-robust potential links, or more formally a path in $\boldsymbol{D}$. These paths constitute the furthest extent that best-response chains extending from nodes 5 and 6 can travel. By definition, the component of $\boldsymbol{D}$ containing nodes 5 and 6 contains all such paths.

The previous example pertains to interventions that perturb non-robust potential links, specifically A_{56}. Interventions that perturb robust links, such as A_{67}, trigger best-response dynamics in all strategic neighborhoods containing the link, which would be the two right-most strategic neighborhoods in Figure 1.

These examples demonstrate that the sizes of components C_i and strategic neighborhoods C_i^+ (since they contain the components) are indicative of the degree of cross-sectional dependence. The key idea is that *if strategic neighborhoods are relatively small, then best-response chains cannot extend too far from the initial perturbation, corresponding to weaker dependence.* To control the length of best-response chains, in §4.3, we state a "subcriticality" condition that ensures that component sizes are asymptotically bounded. From the discussion following Example 7, this should hold if strategic interactions are sufficiently weak.

To derive the condition, we employ a well-known technique used in random graph theory for bounding the size of a component, which is to construct a branching process (see §SA.2.1 for a formal definition) whose size stochastically dominates that of the component (e.g. Bollobás and Riordan, 2012). The basic idea is to explore each C_i via breadth-first search by starting at i, branching to its neighbors, neighbors of its neighbors, and so on. This is akin to growing a branching process, a model of population growth in which individuals in a given generation independently produce a random number of offspring, which corresponds to a node's neighborhood size. Subcriticality ensures that the average number of offspring is less than one, in which case the size of the process does not diverge and $|C_i|$ is asymptotically bounded.

While this is enough to establish an LLN (Leung, 2019b), a CLT additionally requires the distribution of $|C_i|$ to have exponential tails. We utilize a tail bound for subcritical branching processes to obtain the desired result; see Lemma SA.2.3 which is proven by Leung (2020) using an argument due to Turova (2012). This is a key ingredient for verifying Assumption 5.

4.2.2 Coordination

The second source of dependence is due to equilibrium selection. Let us shut down the first source of dependence by supposing the perturbation to ζ_{56} is small enough not to change the sign of V_{56}, so there is no change to the network at any point in the best-response dynamics outlined above. However, if equilibrium selection is governed by a mechanism different from best-response dynamics, the perturbation may still substantially affect the network structure.

Suppose $|\mathcal{E}(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)| = 2$. Since the perturbation does not change V_{56}, the set of equilibria is identical before and after the perturbation. However, one can construct a selection mechanism $\lambda_n(\cdot)$ that outputs one equilibrium under the structural primitives before the perturbation but outputs the other after the perturbation. Hence, the perturbation can alter potential links involving nodes external to C_5 even though any best-response chain must be limited to this component. This illustrates how, *under unrestricted equilibrium selection, all nodes may coordinate on the same "signal" (ζ_{56}), resulting in strongly dependent potential links whose realizations all depend on this random variable.*

In §4.4, we state a condition that ensures coordination is "decentralized" in that strategic neighborhoods "separately select" their pairwise stable subnetworks. A consequence is that any perturbation to a node pair's types or random-utility shocks only affects the selection of the equilibrium subnetwork on strategic neighborhoods involving the pair. For example, in Figure 1, it will be the case that a perturbation to ζ_{56} only changes the equilibrium subnetwork on $\{2, \dots, 7\}$ since no other neighborhood contains $\{5, 6\}$. Combined with the subcriticality condition that ensures strategic neighborhood sizes are asymptotically bounded, this implies that the perturbation can only shift equilibrium selection within a bounded subset of nodes, ensuring weak cross-sectional dependence. This will allow us to construct a radius of stabilization that is asymptotically bounded.

4.3 Strength of Interactions

We next state a condition that controls the sizes of strategic neighborhoods, and hence length of best-response chains described in §4.2.1, by restricting the magnitude of strategic interactions. We measure strategic interaction strength by

$$\begin{aligned} p_{r_n}(X_i, Z_i, X_j, Z_j) = \mathbf{P}\big(\sup_s V(r_n^{-1}\|X_i - X_j\|, s, Z_i, Z_j, \zeta_{ij}) > 0 \mid X_i, Z_i, X_j, Z_j\big) \\ - \mathbf{P}\big(\inf_s V(r_n^{-1}\|X_i - X_j\|, s, Z_i, Z_j, \zeta_{ij}) > 0 \mid X_i, Z_i, X_j, Z_j\big). \quad (17) \end{aligned}$$

This is the effect on link formation of changing S_{ij} from its "lowest" to its "highest" possible value, conditional on types. In other words, it is the maximal change in linking probability induced by the strategic component of $V(\cdot)$.

Let $\Phi_z(\cdot \,|\, x)$ be the conditional distribution of Z_i given $X_i = x$. Recall that f is the density of X_1, $\mathcal{T} = \text{supp}(X_1, Z_1)$, and d_z is the dimension of Z_i. For any

$h\colon \mathbb{R}^d \times \mathbb{R}^{d_z} \to \mathbb{R}$, define the mixed norm

$$\|h\|_{\mathbf{m}} = \sup_{x\in\mathbb{R}^d} \left(\int_{\mathbb{R}^{d_z}} h(x,z)^2 \, \mathrm{d}\Phi^*(z) \right)^{1/2},$$

where Φ^* is a measure defined in the next assumption. Finally, for $\bar{f} = \sup_{x\in\mathbb{R}^d} f(x)$ and κ in (5), let

$$h_{r_n}(x,z) = n\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^{d_z}} p_{r_n}(x,z,x',z')^2 \, \mathrm{d}\Phi_z(z' \mid x') \right)^{1/2} f(x')\,\mathrm{d}x' \quad \text{and}$$

$$h^*(x,z) = \kappa \bar{f} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^{d_z}} p_1(x,z;x',z')^2 \, \mathrm{d}\Phi^*(z') \right)^{1/2} \mathrm{d}x'.$$

Assumption 7 (Subcriticality)**.** *(a) There exists a measure Φ^* on $\mathbb{R}^{d_z}$ such that for all $(x,z) \in \mathcal{T}$ and $n \in \mathbb{N}$,*

$$n\int_{\mathbb{R}^d}\int_{\mathbb{R}^{d_z}} p_{r_n}(x,z;x',z')\, d\Phi_z(z' \mid x') f(x')\, dx' \leqslant \kappa\bar{f} \int_{\mathbb{R}^d}\int_{\mathbb{R}^{d_z}} p_1(x,z;x',z')\, d\Phi^*(z')\, dx',$$

and $h_{r_n}(x,z) \leqslant h^*(x,z)$. (b) $\|h^*\|_{\mathbf{m}} < 1$.

Part (a) is a regularity condition corresponding to Assumption 2 of Leung (2020) that replaces $\Phi_z(\cdot \mid x')$ with a "dominating" measure Φ^*. The substantive requirement is (b), which is a slightly stronger version of Assumption 6 of Leung (2019b). Its formulation is as primitive as possible at this level of generality, but we may obtain further insight by specializing to a particular joint surplus function. The next example demonstrates how the condition restricts the magnitude of strategic interactions.

Example 9. Consider Example 1, and suppose $X_i \sim \mathcal{U}([0,1]^2)$ and $\zeta_{ij} \sim \mathcal{N}(0,1)$. By a change of variables $x' \mapsto x + r_n(x'-x)$ as in (6),

$$\begin{aligned} n&\int_{\mathbb{R}^d}\int_{\mathbb{R}^{d_z}} p_{r_n}(x,z;x',z')\,\mathrm{d}\Phi_z(z' \mid x') f(x')\,\mathrm{d}x' \\ &= n\int_{\mathbb{R}^d} \mathbf{P}\big(-\theta_2 < \theta_1 - \theta_3(r_n^{-1}\|x-x'\|)^2 + \zeta_{12} \leqslant 0\big)\mathbf{1}\{x' \in [0,1]^2\}\,\mathrm{d}x' \\ &\leqslant \kappa \int_{\mathbb{R}^d} \mathbf{P}\big(-\theta_2 < \theta_1 - \theta_3\|x-x'\|^2 + \zeta_{12} \leqslant 0\big)\,\mathrm{d}x'. \end{aligned}$$

Notice the second line equals $h_{r_n}(x, z)$ and the third line $h^*(x, z)$. Letting $\Phi(\cdot)$ denote the normal CDF,

$$\begin{aligned} h^*(x, z) &= \kappa \int_{\mathbb{R}^2} \left[\Phi(\theta_1 + \theta_2 - \theta_3\|u\|^2) - \Phi(\theta_1 - \theta_3\|u\|^2)\right] \mathrm{d}u \\ &= 2\kappa\pi \int_0^\infty w \left[\Phi(\theta_1 + \theta_2 - \theta_3 w^2) - \Phi(\theta_1 - \theta_3 w^2)\right] \mathrm{d}w \end{aligned}$$

by a change of variables to polar coordinates, where π is the transcendental constant. Let $q(x) = x\Phi(x) + \phi(x)$, where $\phi(\cdot)$ is the normal PDF. The integral has a closed-form expression

$$h^*(x, z) = \frac{\kappa\pi}{\theta_3}\big(q(\theta_1 + \theta_2) - q(\theta_1)\big).$$

Assumption 7 requires this to be less than one, which holds if θ_2 is sufficiently small in magnitude. This is transparently a restriction on the strength of strategic interactions.

The example illustrates how $\|h^*\|_{\mathbf{m}} < 1$ *constitutes the network formation analog of well-known weak dependence conditions for linear spatial or temporal autoregressive models*, which require the magnitude of the autoregressive coefficient to be bounded below one. For instance, in the linear-in-means model of peer effects, it is assumed that the endogenous peer effect satisfies $|\beta| < 1$ (Bramoullé et al., 2009). For nonlinear models such as ours, the analogous condition is necessarily more complicated to state. Equation (9) of de Jong and Woutersen (2011) states the condition for dynamic binary choice time series models, which shares some visual similarities with our assumption.

The connection between Assumption 7 and the discussion in §4.2 regarding the size of C_i is that (17) $= \mathbf{E}[D_{ij} \mid X_i, Z_i, X_j, Z_j]$, so by Jensen's inequality and the change of variables argument in (6), $h^*(x, z) \geqslant \mathbf{E}[\sum_j D_{ij} \mid X_i = x, Z_i = z]$. The expectation of the right-hand side is the expected degree in $\boldsymbol{D}$, so by Assumption 7, this is upper bounded by one in expectation. This implies that, as we explore C_i by branching from i to its neighbors in $\boldsymbol{D}$, neighbors of neighbors, and so on, each node has fewer than one neighbor on average. That is, a node at each step is typically replaced by fewer than one node in the next step, so the process is below the replacement rate, and $|C_i|$ should be asymptotically bounded.

4.4 Coordination

The second weak dependence condition restricts the selection mechanism, ruling out coordination of the type described in §4.2.2. It requires selection to be "decentralized" in that each strategic neighborhood C^+ selects its pairwise stable subnetwork based on the types and random-utility shocks of nodes in C^+ alone. In this sense, there is no coordination across disjoint strategic neighborhoods.

Before stating the assumption, we need to clarify why such a restriction is possible. Initially, it may appear incoherent because the pairwise stability of a subnetwork on $H \subseteq \mathcal{N}_n$ can depend on the state of the network outside of H. If nodes $\mathcal{N}_n \backslash H$ are removed from the model, we would expect nodes in H to adjust their links in best response. This suggests that coordination across subsets of nodes is generally unavoidable.

However, if H is specifically a strategic neighborhood, then under Assumption 2, its subnetwork is in fact pairwise stable regardless of the state of the ambient network. Consider Figure 1 and a counterfactual intervention that removes all nodes from the network outside of $C_5^+ = \{2, \dots, 7\}$. The intervention has no impact on the pairwise stability of A_{23} and A_{67} since these are robust links. It also has no impact on the pairwise stability of non-robust potential links between nodes in C_5^+, for example A_{45} since neither nodes 4 nor 5 are linked to $\mathcal{N}_n \backslash C_5^+$ in $\boldsymbol{A}$ and strategic interactions are local by Assumption 2. That is, the non-robust potential links in C_5^+ are separated from $\mathcal{N}_n \backslash C_5^+$ by a "buffer" of robust links, so the intervention does not trigger a chain of best-responses that affects $\boldsymbol{A}_{C_5^+}$. This illustrates the following result.

Proposition 1 (Leung (2019b), Proposition 1). *Under Assumption 2, for any $r > 0$, $n \in \mathbb{N}$, and strategic neighborhood C^+,*

$$\mathcal{E}(r^{-1}\boldsymbol{T}_{C^+}, \boldsymbol{\zeta}_{C^+}) = \{\boldsymbol{A}_{C^+} : \boldsymbol{A} \in \mathcal{E}(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)\}.$$

The left-hand side is the set of pairwise stable networks under the counterfactual model in which the set of nodes is C^+, rather than $\mathcal{N}_n$. The right-hand side takes the set of pairwise stable networks on $\mathcal{N}_n$ and restricts them to C^+. The proposition asserts the two sets are equivalent, which says that the pairwise stability of subnetworks on C^+ only depends on the structural primitives of nodes in C^+. We emphasize that this property is unique to strategic neighborhoods and is not true for arbitrary subsets of nodes. Leung (2020) exploits Proposition 1 to devise an algorithm that computes $\mathcal{E}(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$ in polynomial time under Assumption 7. We utilize it to establish a CLT.

Recall that our objective is to impose the assumption that strategic neighborhoods "separately select" their own pairwise stable subnetworks. Proposition 1 ensures that it is coherent to refer to the pairwise stability of a subnetwork on a strategic neighborhood in isolation from the rest of the network. However, a second concern is that strategic neighborhoods do not necessarily partition $\mathcal{N}_n$, so such an assumption may still appear incoherent. For instance, in Figure 1, the left and middle strategic neighborhoods share nodes 2 and 3 in common, so it is not clear how the two neighborhoods can separately select equilibria.

Recall that components of $\boldsymbol{D}$ do partition $\mathcal{N}_n$, while strategic neighborhoods are obtained by adding nodes that are robustly linked to components. Then *the link between any node pair that lies in multiple strategic neighborhoods must necessarily be robust and therefore have the same realization under any pairwise stable equilibrium.* That is, any pairwise stable subnetwork on the left strategic neighborhood in Figure 1 sets $A_{23} = 1$, as does any pairwise stable subnetwork on the middle strategic neighborhood. Consequently, it is a logically coherent operation to select an equilibrium subnetwork for each strategic neighborhood and then take a "union" to obtain the overall network $\boldsymbol{A}$ (see §SA.4.2 of Leung, 2020, for a detailed elaboration of this idea).

We are now prepared to state the assumption. Let $\lambda_n(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)\big|_H$ be the restriction of the range of $\lambda_n(\cdot)$ to subnetworks on H. For example, under Assumption 3, $\lambda_n(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n) = \boldsymbol{A}$, so $\lambda_n(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)\big|_H = \boldsymbol{A}_H$.

Assumption 8 (Decentralized Selection)**.** *For any $r > 0$, $n \in \mathbb{N}$, and strategic neighborhood C^+ constructed under the structural primitives $(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$, we have $\lambda_n(r^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)\big|_{C^+} = \lambda_{|C^+|}(r^{-1}\boldsymbol{T}_{C^+}, \boldsymbol{\zeta}_{C^+})$.*

This corresponds to Assumption 7 of Leung (2019b). It is important to understand the difference between the left- and right-hand sides of the equality. On the left, we have the model involving all nodes $\mathcal{N}_n$; the selection mechanism produces a network $\boldsymbol{A}$, and we take its subnetwork $\boldsymbol{A}_{C^+}$. On the right, we have the counterfactual model involving only nodes in C^+, and the selection mechanism $\lambda_{|C^+|}(\cdot)$ produces a network

$\lambda_{|C^+|}(r^{-1}\boldsymbol{T}_{C^+}, \boldsymbol{\zeta}_{C^+})$. The assumption asserts that the two outputs are the same.

Sheng (2020) does not require this assumption because she considers a setting with many small independent networks, which means equilibrium selection is necessarily independent across network observations. Assumption 8 is the single-network analog of this requirement, requiring selection to operate separately across latent strategic neighborhoods.

The assumption rules out selection mechanisms in which all nodes coordinate through a common signal, such as a single node's type. Coordination is only allowed to occur within strategic neighborhoods. In the special case where there exists a unique equilibrium on $\mathcal{N}_n$, for instance if there are no strategic interactions, the assumption holds trivially. More generally, the condition is satisfied by variants of myopic best-response dynamics, which are widely used in the theoretical and econometric literature on dynamic network formation (e.g. Jackson, 2010; Mele, 2017).

Example 10 (Best-Response Dynamics)**.** An example of myopic best-response dynamics is the following. Arbitrarily order all node pairs and begin at an arbitrary network $\boldsymbol{A}_0$. At step t, update the previous network $\boldsymbol{A}_{t-1}$ by setting the ijth component to $\mathbf{1}\{V(r_n^{-1}\|X_i - X_j\|, S_{ij}^{t-1}, Z_i, Z_j, \zeta_{ij}) > 0\}$, where $S_{ij}^{t-1} = S_n(i, j, r_n^{-1}\boldsymbol{T}_n, \boldsymbol{A}_{t-1})$, the network statistics evaluated at the prior network. Repeat for all pairs of nodes to obtain $\boldsymbol{A}_t$. Repeat this process until convergence to a network $\boldsymbol{A} \in \mathcal{E}(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$.[7] This constitutes a selection mechanism $\lambda_n(\cdot)$ since it maps structural primitives to an equilibrium network. To ensure that $\lambda_n(\cdot)$ is permutation-equivariant (Assumption 3), we may suppose that potential links in $\boldsymbol{A}_0$ and the ordering of node pairs are only functions of the types of the nodes in the pair.[8]

To understand how this satisfies Assumption 8, first consider the ideal scenario in which the data consists of two independent network observations formed by these dynamics. The equilibrium on each network may be generated in two equivalent ways. First, the dynamics may be run separately on each network. Second, the two networks may be concatenated into one, defining the joint surplus between pairs of nodes in different networks as $-\infty$, and the dynamics may be run on the entire entity.

[7]The usual method for proving the existence of a pairwise stable network is to establish non-existence of "closed cycles," which implies that myopic best-response dynamics always converge to an equilibrium. See for example Proposition 2.1 of Sheng (2020).

[8]More generally, the initial network and ordering of node pairs may be determined by any permutation-equivariant functions of $(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)$.

These produce the same output given the initial network and node pair ordering.

Now suppose the data consists of a single network comprised of two strategic neighborhoods. If the neighborhoods do not share a pair of nodes in common, then this is the same situation as the two-network case. If they do, then both ways of running the dynamics still produce the same result because any pair of nodes shared by both neighborhoods must form a robust link. This reasoning immediately extends to an arbitrary number of strategic neighborhoods, so Assumption 8 holds.

4.5 Main Results

The last assumption we require is a regularity condition.

Assumption 9 (Regularity)**.** *Either* $p_r(X_1, Z_1; X_2, Z_2) = 0$ *a.s. for any* $r > 0$ *in a neighborhood of zero, or* $\inf\{\liminf_{n\to\infty} n\,\mathbf{E}[p_{r_n}(x, z; X_2, Z_2)] \colon (x, z) \in \mathcal{T}\} > 0$.

The case $p_r(X_1, Z_1; X_2, Z_2) = 0$ corresponds to a model without strategic interactions, which is only mentioned for completeness. In the more interesting case, the assumption essentially requires that strategic interactions are sufficiently nontrivial for all nodes in that $p_{r_n}(X_1, Z_1; X_2, Z_2)$ is at least order n^{-1}. This is a mild requirement that is typically satisfied because the two probabilities in (17) are upper and lower bounds on the probability of link formation, and the upper bound is order n^{-1} under sparsity by (6).

Example 11. Consider Example 9. Following the derivation there,

$$n\,\mathbf{E}[p_r(x, z; X_2, Z_2)] = 2nr^2\pi \int_0^\infty w\left[\Phi(\theta_1 + \theta_2 - \theta_3 w^2) - \Phi(\theta_1 - \theta_3 w^2)\right] \mathrm{d}w.$$

If $\theta_2 = 0$, this corresponds to the case $p_r(X_1, Z_1; X_2, Z_2) = 0$ a.s. Otherwise, Assumption 9 holds if $\theta_2 \neq 0$ since $nr_n^2 = \kappa > 0$ by (5).

Theorem 2. *Assumptions 1–4 and 7–9 imply Assumption 5.*

Proof. See §SA.2. We sketch the proof below. ∎

From Theorems 1 and 2 we immediately obtain the following CLT for strategic network formation.

Corollary 1. *Under Assumptions 1–4 and 6–9,* $\sup_n \|\mathbf{\Sigma}_n\|_\infty < \infty$*, and if additionally* $\liminf_{n\to\infty} \lambda_{min}(\mathbf{\Sigma}_n) > 0$*, then* (12) *holds.*

At this level of generality, these conditions are close to as primitive as possible, but with additional structure we can derive lower-level conditions.

Corollary 2. *Consider the network formation model in Example 9, and assume Assumption 6(b) and the following weak dependence conditions hold.*

(a) (Strength of interactions) For $q(x)$ *defined in the example,*

$$\frac{\kappa\pi}{\theta_3}\big(q(\theta_1+\theta_2) - q(\theta_1)\big) < 1.$$

(b) (Decentralized selection) The equilibrium is selected via myopic best-response dynamics as in Example 10.

If $\sum_{i=1}^n \psi_i(\mathcal{N}_n)$ *is a vector of connected subnetwork counts (Example 5), then* (12) *holds.*

Proof. We verify the conditions of Corollary 1. Assumption 1 holds because

$$\bar{\Phi}_\zeta(\bar{V}^{-1}(\delta,0)) = \bar{\Phi}_\zeta(\theta_3\delta^2 - \theta_1 - \max\{\theta_2, 0\}),$$

which decays to zero exponentially with δ since $\bar{\Phi}_\zeta$ is the complementary CDF of the standard normal distribution. Assumption 2 holds by choice of S_{ij}. Assumptions 3 and 8 hold by (b). Assumption 4 holds because we consider connected subnetwork counts (see Example 6). Assumption 6(a) follows from Proposition A.1. Assumption 7 follows from (a) (see Example 9). Finally, Assumption 9 holds, as shown in Example 11. ■

4.6 Method of Proof

We present the method of proof for Theorem 2 for the case of the average degree, whose node statistic is $\psi_i(\mathcal{N}_n) = \sum_{j=1}^n A_{ij}$. The approach can be applied to other network models with strategic interactions.[9] We construct an upper bound on the radius of stabilization that has a distribution with sufficiently thin tails.

Step 1. Recalling (9), we construct a set of nodes $J_i \subseteq \mathcal{N}_n$ positioned near i such that

$$\psi_i(\mathcal{N}_n) = \psi_i(J_i). \tag{18}$$

That is, i's node statistic is invariant to the counterfactual removal of nodes outside of J_i. The challenge is to find a set that is relatively small so that the distribution of the set's size has thin tails. Supposing such a set could be found, the radius of stabilization $\boldsymbol{R}_i(n, r_n)$ (Definition 1) would be upper bounded by the radius of the smallest ball centered at $r_n^{-1}X_i$ containing the positions of J_i (plus one):

$$\boldsymbol{R}_i(n, r_n) \leqslant \tilde{\boldsymbol{R}}_i(n, r_n) \equiv \max_{j \in J_i} r_n^{-1} \|X_i - X_j\| + 1. \tag{19}$$

Since $\psi_i(\mathcal{N}_n) = \sum_{j=1}^n A_{ij}$ is simply the 1-neighborhood size, an initial guess for J_i might be i's 1-neighborhood. However, this does not generally satisfy (18). To see why, consider node $i = 6$ in Figure 1 whose 1-neighborhood is contained in $\{4, 5, 7\}$. Since $D_{56} = 1$, the potential link A_{56} is non-robust and therefore may differ under the counterfactual that removes node 3 from the model, even though 3 is not i's neighbor.

Instead we take $J_i = C_i^+$. The pairwise stability of $\boldsymbol{A}_{C_i^+}$ is invariant to the removal of nodes outside of C_i^+ (see Example 8 and Proposition 1), and under Assumption 8, $\boldsymbol{A}_{C_i^+}$ remains the selected equilibrium subnetwork on C_i^+ after removal of $\mathcal{N}_n \backslash C_i^+$. That is, for $H = C_i^+$, $\boldsymbol{A}_H = \lambda_n(r_n^{-1}\boldsymbol{T}_n, \boldsymbol{\zeta}_n)|_H = \lambda_{|H|}(r_n^{-1}\boldsymbol{T}_H, \boldsymbol{\zeta}_H)$. Since C_i^+ contains i's 1-neighborhood, (18) holds.

Step 2. Next we show that $|J_i|$ has a distribution with exponential tails. First, §SA.2.1 establishes that $|J_i|$ is stochastically dominated by the size of a certain branching process when J_i is constructed from components and K-neighborhoods of dyadic networks. Lemma SA.2.6 shows that the size of the branching process has a distribution with exponential tails using Assumptions 7 and 9.

In the case of average degree, our choice of $J_i = C_i^+$ is the union of a component of $\boldsymbol{D}$ and 1-neighborhoods in $\boldsymbol{\Pi}$, as defined in §4, which are dyadic networks. For more general K-neighborhood node statistics satisfying Assumption 4, the construction of

[9] A previous version of the paper presents applications to dynamic network formation (Leung and Moon, 2019), and Leung (2019a) applies the methodology to games on networks.

J_i is more complicated but still involves similar objects. Lemma SA.2.1 in §SA.2.2 shows how to construct J_i for a general node statistic satisfying Assumption 4. The lemmas in §SA.2.2 can then be applied to obtain the desired tail bounds.

Step 3. We translate the tail bound for $|J_i|$ into one for $\tilde{\boldsymbol{R}}_i(n, r_n)$. Intuitively, if $|J_i|$ is small, then so will be $\tilde{\boldsymbol{R}}_i(n, r_n)$ since nodes are homophilous in positions (Assumption 1), so each $j \in J_i$ will typically be close to i in terms of distance $r_n^{-1}\|X_i - X_j\|$. Lemma SA.2.8 provides the formal argument.

5 Applications to Inference

We discuss two inference procedures applicable to network data generated by the model in §2. Our results provide the first formal justification for their use in the subsequent applications.

Since our objective is to establish a CLT, thus far we have made no distinction between what aspects of the model are known or observed by the econometrician. In what follows, the only requirement is that the econometrician must be able to compute the relevant network moments, but it is application-specific which structural primitives need to be observed for this to be possible.

In the network statistics applications, only $\boldsymbol{A}$ needs to be observed. In structural applications, $V(\cdot)$ is typically known up to a vector of parameters, $\boldsymbol{A}$ is observed, and ζ_{ij} is unobserved, but any subvector of (X_i, Z_i) could potentially be unobserved. In particular, the procedures can be implemented when positions are unobserved, as in the literature on latent-space models. Typically the distribution of the unobserved component of types, conditional on observables, is assumed known up to a vector of parameters.

5.1 Network Statistics

Define $\mu_0 = \mathbf{E}[\psi_1(\mathcal{N}_n)]$ and $\hat{\mu} = n^{-1}\sum_{i=1}^n \psi_i(\mathcal{N}_n)$. Consider testing the null hypothesis

$$H_0\colon \mu_0 = \mu.$$

This is relevant for the reporting of stylized facts in the networks literature. Such facts are obtained by computing various statistics from $\boldsymbol{A}$ but are seldom accompanied by formal uncertainty quantification due to a lack of available methods. Leung (2022) discusses two particular examples: testing for nontrivial clustering (his §3.3) and testing for a power law degree distribution (his §3.4). We next discuss two generic tests justified by our CLT. Confidence regions for μ_0 can be obtained by test inversion.

Single large network. If the sample consists of a single network, we may apply the resampling procedure due to Song (2016) and Leung (2022). Let α be the desired level of the test, $R_n = (n/2)^{4/3}$ rounded to the nearest integer, Π be the set of all bijections (permutation functions) on $\{1, \dots, n\}$, and $\pi = (\pi_r)_{r=1}^{R_n}$ be i.i.d. uniform draws from Π. Let $\hat{\boldsymbol{V}} = n^{-1} \sum_{i=1}^n (\psi_i(\mathcal{N}_n) - \hat{\mu})(\psi_i(\mathcal{N}_n) - \hat{\mu})'$, the sample variance. Define the test statistic

$$T_U(\mu; \pi) = \frac{1}{\sqrt{d_\psi R_n}} \sum_{r=1}^{R_n} (\psi_{\pi_r(1)}(\mathcal{N}_n) - \mu)' \hat{\boldsymbol{V}}^{-1} (\psi_{\pi_r(2)}(\mathcal{N}_n) - \mu),$$

recalling that d_ψ is the dimension of the range of $\psi_n(\cdot)$. Let $z_{1-\alpha}$ be the $1-\alpha$ quantile of the standard normal distribution. The test rejects if and only if

$$T_U(\mu; \pi) > z_{1-\alpha}. \tag{20}$$

Validity of this test hinges on the high-level weak dependence condition that $\hat{\mu}$ is $\sqrt{n}$-consistent. Leung (2019b) provides an LLN for $\hat{\mu}$ but not a rate of convergence. Our paper is the first to provide primitive conditions for $\sqrt{n}$-consistency in the context of strategic network formation.

For intuition on the importance of $\sqrt{n}$-consistency for the test's validity, consider a simpler test statistic studied by Leung (2022):

$$\tilde{T}_M(\mu; \pi) = \frac{1}{\sqrt{R_n^M}} \sum_{r=1}^{R_n^M} \hat{\boldsymbol{V}}^{-1/2} \big(\psi_{\pi_r(1)}(\mathcal{N}_n) - \mu\big)$$

for $R_n^M = \sqrt{n}$ rounded to the nearest integer. To understand its asymptotic behavior, we add and subtract its mean conditional on the data:

$$\frac{1}{\sqrt{R_n^M}}\sum_{r=1}^{R_n^M}\hat{\boldsymbol{V}}^{-1/2}\big(\psi_{\pi_r(1)}(\mathcal{N}_n)-\mathbf{E}[\psi_{\pi_r(1)}(\mathcal{N}_n)\mid\{\psi_i(\mathcal{N}_n)\}_{i=1}^n]\big)$$
$$+\frac{1}{\sqrt{R_n^M}}\sum_{r=1}^{R_n^M}\hat{\boldsymbol{V}}^{-1/2}\big(\mathbf{E}[\psi_{\pi_r(1)}(\mathcal{N}_n)\mid\{\psi_i(\mathcal{N}_n)\}_{i=1}^n]-\mu\big).$$

Conditional on the data $\{\psi_i(\mathcal{N}_n)\}_{i=1}^n$, the permutations π_r are independent, so the first term is an average of R_n^M conditionally independent observations and can be shown to be asymptotically $\mathcal{N}(\mathbf{0},\boldsymbol{I})$ under H_0. The second is a bias term that can be shown to equal $(R_n^M/n)^{1/2}\hat{\boldsymbol{V}}^{-1/2}\sqrt{n}(\hat{\mu}-\mu)$. Hence, under H_0 and $\sqrt{n}$-consistency, the bias is order $(R_n^M/n)^{1/2}=o(1)$.

Multiple large networks. A drawback of the previous procedure is that it is inefficient, having a rate of convergence slower than $\sqrt{n}$ since $R_n^M=o(n)$. If the sample consists of sufficiently many independent large networks, then more powerful methods are available from the cluster-robust inference literature. Consider a sequence of L independent networks indexed by n, where L is fixed with respect to n and each network $\ell=1,\dots,L$ has size n_ℓ satisfying $n_\ell/n\to c_\ell\in(0,\infty)$ as $n\to\infty$. Let $\hat{\mu}^\ell=n_\ell^{-1}\sum_{i=1}^{n_\ell}\psi_i(\mathcal{N}_{n_\ell})$, the network moment computed on network ℓ. Assume there exists a universal population moment μ_0 such that $\mathbf{E}[\hat{\mu}^\ell]=\mu_0+o(n^{-1/2})$ for all ℓ. We seek to test the null $H_0\colon \mu_0=\mu$.

We consider the randomization test proposed by Canay et al. (2017). For $S_{n,\ell}=\sqrt{n_\ell}(\hat{\mu}^\ell-\mu)$ and $S_n=(S_{n,\ell})_{\ell=1}^L$, define the Wald statistic

$$T(S_n)=\left(\frac{1}{\sqrt{L}}\sum_{\ell=1}^L S'_{n,\ell}\right)\left(\frac{1}{L}\sum_{\ell=1}^L S_{n,\ell}S'_{n,\ell}\right)^{-1}\left(\frac{1}{\sqrt{L}}\sum_{\ell=1}^L S_{n,\ell}\right).$$

We obtain critical values from the randomization distribution $\{T(\pi S_n)\colon \pi\in\{-1,1\}^L\}$ where $\pi S_n=(\pi_\ell S_{n,\ell})_{\ell=1}^L$ for $\pi=(\pi_\ell)_{\ell=1}^L$. Let α be the desired level of the test, $q=2^L(1-\alpha)$ rounded up to the nearest integer, and $c_{L,1-\alpha}$ be the qth largest value of $\{T(\pi S_n)\colon \pi\in\{-1,1\}^L\}$. The test rejects if and only if

$$T(S_n)>c_{L,1-\alpha}.$$

The test is asymptotically level α under the high-level condition that the limit distribution of the vector of network moments is asymptotically normal. We provide the first primitive sufficient conditions in the literature for strategic network formation. Intuitively, under H_0 and asymptotic normality, $\{S_{n,\ell}\}_{\ell=1}^L$ are independent draws from a mean-zero, approximately normal distribution. Hence, multiplying these draws by ± 1 does not change the asymptotic distribution of $T(S_n)$, which is the key justification for the validity of randomization tests.

5.2 Structural Inference

We revisit Example 6, which concerns inference on structural parameters using moment inequalities proposed by Sheng (2020). Recalling the setup there, let θ_0 denote the true parameters of $V(\cdot)$. To test the hypothesis $H_0\colon \theta_0 = \theta$, we test the moment inequality

$$\mu_0 \equiv \mathbf{E}[\psi_1(\mathcal{N}_n)] \leqslant \mathbf{0} \quad \text{where} \quad \psi_{i_1}(\mathcal{N}_n) = \sum_{i_2=1}^{n} \dots \sum_{i_m=1}^{n} (G_{i_1,\dots,i_m} - H_{i_1,\dots,i_m}(\theta_0)) q_{i_1,\dots,i_m}(\boldsymbol{T}_n),$$

When the data consists of a single large network, we can employ a test due to Leung (2022). Let μ_k be the kth component of μ_0, $\psi_{ik}(\mathcal{N}_n)$ the kth component of $\psi_i(\mathcal{N}_n)$, and $T_{U,k}(\mu_k;\pi)$ the U-type statistic defined in the previous subsection but computed with scalar data $\{\psi_{ik}(\mathcal{N}_n)\}_{i=1}^n$. Also let $\hat{\mu}_k$ be the kth component of $\hat{\mu}$ and $\hat{\boldsymbol{V}}_{kk}$ the kth diagonal entry of $\hat{\boldsymbol{V}}$. Define the test statistic

$$Q_n(\pi) = \max_{1 \leqslant k \leqslant d_\psi} \left\{ T_{U,k}(0;\pi) - \mu_k^* \mathbf{1}\{\hat{\mu}_k < 0\} \right\} \quad \text{where}$$
$$\mu_k^* = \hat{\mu}_k \hat{\boldsymbol{V}}_{kk}^{-1} \frac{1}{\sqrt{d_\psi R_n}} \sum_{r=1}^{R_n} \left(\psi_{\pi_r(1),k}(\mathcal{N}_n) + \psi_{\pi_r(2),k}(\mathcal{N}_n) \right) - \sqrt{\frac{R_n}{d_\psi}} \hat{\boldsymbol{V}}_{kk}^{-1} \hat{\mu}_k^2.$$

Let $\tilde{\pi}_1, \dots, \tilde{\pi}_L$ be i.i.d. with the same distribution as π. Let $q = L(1-\alpha)$ rounded up to the nearest integer and $c_{L,1-\alpha}$ be qth largest value of $\{\max_{1 \leqslant k \leqslant d_\psi} T_{U,k}(\hat{\mu}_k; \tilde{\pi}_\ell)\}_{\ell=1}^L$. The test rejects if and only if

$$Q_n(\pi) > c_{L,1-\alpha}.$$

Theorem 2 of Leung (2022) provides conditions under which the test is asymptotically level α under H_0. The main assumption that needs to be verified is $\sqrt{n}$-consistency of $\hat{\mu}$, which is a consequence of our CLT. Like the test in the previous subsection, this procedure is inefficient with a slower than $\sqrt{n}$-rate of convergence. To construct more powerful tests, we require either a consistent estimate of $\mathbf{\Sigma}_n$ or a valid resampling procedure, topics we leave to future research.

6 Simulation Study

We conduct a simulation study to assess the quality of the normal approximation and finite-sample performance of the inference procedures in §5. We simulate data according to the model in Example 1 with $\|\cdot\|$ equal to the Euclidean norm, $\theta = (1, 0.25, 1)$, $X_i \overset{iid}{\sim} \mathcal{U}([0,1]^2)$, and $\zeta_{ij} \overset{iid}{\sim} \mathcal{N}(0,1)$. To satisfy Assumption 7, we set $\kappa = 0.8\theta_3\pi^{-1}(q(\theta_1+\theta_2)-q(\theta_1))^{-1}$ following the notation in Example 9, which implies $\|h^*\|_{\mathbf{m}} = 0.8 < 1$. To satisfy Assumption 8, we select the equilibrium network using myopic best-response dynamics (Example 10) starting from the network $\mathbf{\Pi}$ defined in §4.1. The model generates networks with fairly realistic properties; the largest component comprises about 75 percent of the network, while the average degree and clustering are respectively about 4.3 and 0.4.

We consider equality tests involving the following network moment. Define the node statistic $\psi_i(\mathcal{N}_n) = \psi_i^c - \psi_i^f$ where

$$\psi_i^c = \frac{\sum_{j\neq k\neq i} A_{ij}A_{jk}A_{ik}}{\sum_{j\neq k\neq i} A_{ij}A_{ik}} \quad \text{and} \quad \psi_i^f = \frac{2}{n-1}\sum_{j\neq i} A_{ij}.$$

Then $n^{-1}\sum_{i=1}^n \psi_i^c$ is the average clustering coefficient, a measure of triadic closure, while $n^{-1}\sum_{i=1}^n \psi_i^f$ is the link frequency. The motivation for the network moment $n^{-1}\sum_{i=1}^n \psi_i(\mathcal{N}_n)$ is as follows. Under the "null" Erdős-Rényi model, the moment is approximately zero. However, a well-known stylized fact is that real-world networks typically feature nontrivial clustering in that the statistic is far from zero (Barabási, 2015). While this fact is based only on the point estimate, the inference procedures in §5 enable us to formulate this as a statistical test.

We simulate rejection rates for three different tests of hypotheses of the form

$$H_0\colon \mathbf{E}[\psi_1] = c\mu \quad \text{against} \quad H_1\colon \mathbf{E}[\psi_1] \neq c\mu$$

where $c \in (0,1]$ and μ is the true value of $\mathbf{E}[\psi_1]$, computed using 40k simulation draws. To simulate size, we set $c = 1$, and to simulate power, we set $c \in \{0.8, 0.9\}$. We consider the following tests at the 5-percent level. To assess the quality of the normal approximation, we use the "oracle" t-test with test statistic $(n^{-1}\sum_{i=1}^{n}\psi_i - c\mu)/\sigma$, where σ is the true standard error, computed by taking the standard deviation of the network moment across 40k simulation draws. We then consider the dependence-robust test (20) and randomization test from §5.1. The randomization test will use data from 5, 6, or 8 independent networks of identical size, while the other tests only utilize data from one of the networks. We simulate size and power using 5k simulation draws.

Table 1: Simulation results

	Size ($c = 1$)			Power ($c = 0.9$)			Power ($c = 0.8$)		
n	250	500	1000	250	500	1000	250	500	1000
Oracle	0.050	0.054	0.048	0.337	0.627	0.915	0.856	0.995	1.000
DR	0.069	0.057	0.058	0.176	0.263	0.410	0.530	0.823	0.981
Rand (8)	0.040	0.040	0.038	0.928	0.998	1.000	1.000	1.000	1.000
Rand (6)	0.028	0.029	0.030	0.672	0.929	0.997	0.993	1.000	1.000
Rand (5)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
μ	0.393	0.409	0.416	0.393	0.409	0.416	0.393	0.409	0.416
σ	0.026	0.018	0.013	0.026	0.018	0.013	0.026	0.018	0.013

5000 simulations. DR corresponds to the dependence-robust test and Rand (x) to the randomization test using data from x independent networks, all of size n. Oracle corresponds to the t-test using the true standard error σ.

Table 1 shows the results. From the first three columns, we see that the oracle t-test controls size well across all sample sizes, illustrating the quality of the normal approximation. The randomization test (the Rand rows in the table) controls size well across all n, while the dependence-robust test (the DR row in the table) exhibits some over-rejection, particularly in smaller samples. Section 6 of Leung (2022) provides additional simulation results for the dependence-robust test under a variety of data-generating processes, including a network formation model similar to the present design. His results show that the test has good properties for different choices of the tuning parameter R_n.

The last six columns of Table 1 compare power. The differences between the oracle and dependence-robust tests are due to the latter's slower rate of convergence. A comparison of the power of the randomization and dependence-robust tests is not apples-to-apples because the former utilizes 5 to 8 times the sample size. However, we see that the randomization test requires at least 6 networks to obtain nontrivial power. These results confirm the discussion in §5.

7 Conclusion

This paper develops a large-sample theory for a model of network formation with strategic interactions and homophilous agents when the data consists of a small sample of large networks or possibly a single network. We prove a general CLT under a high-level weak dependence condition and provide a general methodology for its verification. We apply the methodology to obtain primitive conditions for a CLT for moments of pairwise stable networks.

There are several important directions for future work. Concentration inequalities for stabilization would be useful, for example, for deriving lower-level conditions for uniform convergence of nonparametric or high-dimensional estimators using network data. It is also of interest to develop efficient alternatives to the inference procedures in §5 in the case where the data is a single network observation.

A Bounded Subnetwork Moments

This section formally defines subnetwork counts, introduced in Example 5, and verifies Assumption 6(a) for these moments. Fix the subnetwork size $h \in \mathbb{N}\backslash\{1\}$, and let a_h be a connected network on nodes $\{1, \dots, h\}$ (a network is connected if the path distance between any pair of nodes is finite). For $H \subset \mathcal{N}_n$ with $|H| = h$, we say $\boldsymbol{A}_H$ is *isomorphic* to a_h if there exists a bijection $\pi\colon \mathcal{N}_n \to \mathcal{N}_n$ such that $\pi(\boldsymbol{A})_H = a_h$, where $\pi(\boldsymbol{A})$ is the permuted adjacency matrix $(A_{\pi(i)\pi(j)})_{i,j \in \mathcal{N}_n}$. If $\boldsymbol{A}_H$ is isomorphic to a_h, we write $\boldsymbol{A}_H \cong a_h$.

Let $[n]_h$ be the set of subsets of $\mathcal{N}_n$ of size h. The subnetwork count for a_h is

$$\sum_{H \in [n]_h} \mathbf{1}\{\boldsymbol{A}_H \cong a_h\}.$$

This counts "unlabeled" subnetworks isomorphic to a_h. To rewrite it as a (scaled) network moment $\sum_{i=1}^n \psi_i(\mathcal{N}_n)$, we observe that this is proportional to the corresponding count of "labeled" subnetworks. Formally,

$$h! \sum_{H \in [n]_h} \mathbf{1}\{\boldsymbol{A}_H \cong a_h\} = \sum_{i_1 \in \mathcal{N}_n} \underbrace{\sum_{i_2 \in \mathcal{N}_n} \dots \sum_{i_h \in \mathcal{N}_n} \mathbf{1}\{\boldsymbol{A}_{\{i_1,\dots,i_h\}} \cong a_h\}}_{\psi_{i_1}(\mathcal{N}_n)}. \tag{A.1}$$

The right-hand side counts labeled subnetworks isomorphic to a_h, which will be our object of analysis in what follows.

Equation (5.3) of Sheng (2020) uses subnetwork counts to define moment inequalities. She divides the counts by $\binom{n}{h}$ since she considers a setting with many small independent networks. With a single large sparse network, the correct scaling is instead to divide by n, resulting in the network moment $n^{-1}\sum_{i=1}^n \psi_i(\mathcal{N}_n)$ for $\psi_i(\mathcal{N}_n)$ defined in (A.1).

Proposition A.1. *Under Assumption 1, Assumption 6(a) holds for node statistics $\psi_i(\mathcal{N}_n)$ of the form given in* (A.1).

Proof. Per the setup of Assumption 6(a), let the number of nodes be $m+k$, where $k \in \{1,2\}$ and m is either a nonrandom element of $\mathbb{N}$ or equal to N_n. The statement of Assumption 6(a) also considers models where the set of nodes is a subset $H_n \cup \{1\}$ of $\mathcal{N}_{m+k}$, but since this only reduces the upper bound in (A.3) below, it is sufficient to consider $H_n = \mathcal{N}_{m+k}\backslash\{1\}$. Define

$$\psi_1(\mathcal{N}_{m+k}) = \sum_{i_1 \in \mathcal{N}_{m+k}} \dots \sum_{i_{h-1} \in \mathcal{N}_{m+k}} \mathbf{1}\{\boldsymbol{A}_{\{1,i_1,\dots,i_{h-1}\}} \cong a_h\}. \tag{A.2}$$

Since a_h is a connected network on $\{1,\dots,h\}$, any node in the network is at most path distance $h-1$ from node 1. Therefore, we can replace occurrences of $\mathcal{N}_{m+k}$ in (A.2) with $\mathcal{N}_{\boldsymbol{A}}(1,h-1)$. With this change, (A.2) is bounded above by $|\mathcal{N}_{\boldsymbol{A}}(1,h-1)|^{h-1}$.

Recall the definition of the network $\boldsymbol{M}$ from (35). Since $A_{ij} \leqslant M_{ij}$, $\boldsymbol{A}$ is a subnetwork of $\boldsymbol{M}$, so

$$|\mathcal{N}_{\boldsymbol{A}}(1,h-1)|^{h-1} \leqslant |\mathcal{N}_{\boldsymbol{M}}(1,h-1)|^{h-1}. \tag{A.3}$$

By Lemma SA.2.2, for m sufficiently large, $|\mathcal{N}_{\boldsymbol{M}}(1,h-1)|$ is stochastically dominated by the size of a branching process $|\mathfrak{X}_{r_n}^M(X_1,Z_1;h-1)|$. By Lemma SA.2.4, the distribution of $|\mathfrak{X}_r^M(x,z;h-1)|$ has exponential tails uniformly in x, z, r. It follows that $|\mathcal{N}_{\boldsymbol{M}}(1,h-1)|^{h-1}$ has uniformly bounded p moments for any $p > 2$. ∎

B Data Availability Statement

The code underlying this research is available on Zenodo at https://doi.org/10.5281/zenodo.17807813.

References

Badev, Anton, "Nash Equilibria on (Un)stable Networks," *Econometrica*, 2021, *89* (3), 1179–1206.

Barabási, A., *Network Science*, Cambridge University Press, 2015.

Barabási, A. and R. Albert, "Emergence of Scaling in Random Networks," *Science*, 1999, *286* (5439), 509–512.

Bickel, P., A. Chen, and E. Levina, "The Method of Moments and Degree Distributions for Network Models," *Annals of Statistics*, 2011, *39* (5), 2280–2301.

Bollobás, B. and O. Riordan, "A Simple Branching Process Approach to the Phase Transition in $G(n, p)$," *Electronic Journal of Combinatorics*, 2012, *19* (4), P21.

Boucher, V. and I. Mourifié, "My Friend Far Far Away: A Random Field Approach to Exponential Random Graph Models," *Econometrics Journal*, 2017, *20* (3), S14–S46.

Bramoullé, Y., H. Djebbari, and B. Fortin, "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

Canay, I., J. Romano, and A. Shaikh, "Randomization Tests Under an Approximate Symmetry Assumption," *Econometrica*, 2017, *85* (3), 1013–1030.

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.

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

Charbonneau, K., "Multiple Fixed Effects in Binary Response Panel Data Models," *The Econometrics Journal*, 2017, *20* (3), S1–S13.

Chen, L. and Q. Shao, "Normal Approximation Under Local Dependence," *Annals of Probability*, 2004, *32* (3), 1985–2028.

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman, "An Empirical Model for Strategic Network Formation," in "The Econometric Analysis of Network Data," Elsevier, 2020, pp. 123–148.

de Jong, Robert M. and Tiemen Woutersen, "Dynamic Time Series Binary Choice," *Econometric Theory*, 2011, *27*, 673–702.

Durlauf, S., "Associational Redistribution: A Defense," *Politics & Society*, 1996, *24* (4), 391–410.

Dzemski, A., "An Empirical Model of Dyadic Link Formation in a Network with Unobserved Heterogeneity," *Review of Economics and Statistics*, 2019, *101* (5), 763–776.

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

Graham, B., "An Econometric Model of Network Formation with Degree Heterogeneity," *Econometrica*, 2017, *85* (1), 1033–1063.

Hoff, P., A. Raftery, and M. Handcock, "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 2002, *97* (460), 1090–1098.

Jackson, M., *Social and Economic Networks*, Princeton University Press, 2010.

_ , **B. Rogers, and Y. Zenou**, "Networks: An Economic Perspective," in "Oxford Handbook of Social Network Analysis" 2021, pp. 535–562.

Jochmans, K., "Semiparametric Analysis of Network Formation," *Journal of Business and Economic Statistics*, 2018, *36* (4), 705–713.

Kojevnikov, D., V. Marmer, and K. Song, "Limit Theorems for Network Dependent Random Variables," *Journal of Econometrics*, 2021, *222* (2), 882–908.

Kuersteiner, G., "Limit Theorems for Data with Network Structure," *arXiv preprint arXiv:1908.02375*, 2019.

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

_ , "Inference in Models of Discrete Choice with Social Interactions Using Network Data," *arXiv preprint arXiv:1911.07106*, 2019.

_ , "A Weak Law for Moments of Pairwise-Stable Networks," *Journal of Econometrics*, 2019, *210* (2), 310–326.

_ , "Equilibrium Computation in Discrete Network Games," *Quantitative Economics*, 2020, *11* (4), 1325–1347.

_ , "Dependence-Robust Inference Using Resampled Statistics," *Journal of Applied Econometrics*, 2022, *37* (2), 270–285.

_ **and R. Moon**, "Normal Approximation in Large Network Models," *arXiv preprint arXiv:1904.11060v1*, 2019.

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

Mele, A., "A Structural Model of Dense Network Formation," *Econometrica*, 2017, *85* (3), 825–850.

_ , "Does School Desegregation Promote Diverse Interactions? An Equilibrium Model of Segregation Within Schools," *American Economic Journal: Economic Policy*, 2020, *12* (2), 228–57.

_ , "A Structural Model of Homophily and Clustering in Social Networks," *Journal of Business and Economic Statistics*, 2022, *40* (3), 1377–1389.

Menzel, K., "Strategic Network Formation with Many Agents," *NYU working paper*, 2024.

Penrose, M., *Random Geometric Graphs*, Oxford University Press, 2003.

_ , "Gaussian Limits for Random Geometric Measures," *Electronic Journal of Probability*, 2007, *12*, 989–1035.

_ **and J. Yukich**, "Weak Laws of Large Numbers in Geometric Probability," *Annals of Applied Probability*, 2003, pp. 277–303.

_ **and** _ , "Normal Approximation in Geometric Probability," *arXiv preprint arXiv:math/0409088*, 2008.

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

Ridder, G. and S. Sheng, "Two-Step Estimation of a Strategic Network Formation Model with Clustering," *arXiv preprint arXiv:2001.03838*, 2022.

Sheng, S., "A Structural Econometric Analysis of Network Formation Games Through Subnetworks," *Econometrica*, 2020, *88* (5), 1829–1858.

Song, K., "Ordering-Free Inference from Locally Dependent Data," *arXiv preprint arXiv:1604.00447*, 2016.

Turova, T., "Asymptotics for the Size of the Largest Component Scaled to "$\log n$" in Inhomogeneous Random Graphs," *Arkiv főr Matematik*, 2012, *51* (2), 371–403.