

Adversarial Structural Estimation on Graphs*

Vít Illichmann[†] and Paolo Zacchia[‡]

February 2026

Abstract

We study structural estimation on networks in the empirically common case of a single large observed graph. We propose an adversarial estimator that minimizes statistical distance between observed and simulated node-specific distributions of local network neighborhoods. The paper provides two theoretical results: population identification via a divergence characterization of the estimation criterion function, and consistency under growing-graph asymptotics with cross-observation dependence. A key contribution is computational. We provide a reproducible estimation workflow that integrates fixed-point simulation, efficient focal-neighborhood data construction, and alternating minimax training with stabilization tools suitable for large-scale runs. The workflow is model-agnostic in a broad class of network structural models and is straightforward to implement with modern software. In benchmark simulations, the procedure scales to large graphs and recovers structural parameters with high precision.

JEL Classification Codes: C15, C21, C45

Keywords: Networks, Adversarial Estimation, Structural Models

Code repository: https://github.com/VitaIll/adversarial_networks

*This paper was previously circulated with the title *Convolutional Peer Effects*; it results from research funded by the ERC-CZ project No. LL2319 financed by the Ministry of Education, Youth and Sport of the Czech Republic (*Ministerstvo školství, mládeže a tělovýchovy České republiky*). All errors and omissions are our own.

[†]CERGE-EI; e-mail: vit.illichmann@gmail.com.

[‡]CERGE-EI, Ca' Foscari University of Venice, and CEPR; e-mail: Paolo.Zacchia@cerge-ei.cz.

1 Introduction

Peer, social and network effects are ubiquitous in economic analysis. In settings such as schooling, labor-market referrals, neighborhood policies, and technology diffusion, a policy shock affects treated agents directly and also affects nearby agents through behavioral responses along network links. In these environments, reduced-form peer-effect coefficients are informative but typically insufficient for policy design, because they do not by themselves recover the behavioral primitives that map into equilibrium outcomes on the observed network (Bramoullé, Djebbari, and Fortin 2020; Graham 2015; de Paula, Rasul, and Sousa 2023). Structural estimation based on an explicit economic model can thus provide substantive policy counterfactuals regarding alternative incentives, information environments, or shocks. However, conventional econometric frameworks for structural estimation (e.g., the Method of Simulated Moments) are known to perform poorly in small samples. Issues of cross-observation dependence that are typical of networked data arguably exacerbate this problem.

To enable estimation of arbitrary structural models on network data, this paper develops an adversarial minimum-distance estimator that builds on Kaji, Manresa, and Pouliot (2023). In particular, we focus our analysis on a *single-graph environment*: a single large observed network and an associated equilibrium outcome vector. This corresponds to many real-world networks of interest, and poses nontrivial econometric challenges due to multiple sources of cross-node dependence (exogenous and endogenous, i.e., induced by the structural model in equilibrium). Our contribution is both theoretical and computational. On the theoretical side, we provide both a population identification theorem for an adversarial criterion defined on local subgraphs, and a consistency theorem under growing-graph asymptotics with endogenous dependence. On the computational side, we adapt and enrich the adversarial algorithm by Kaji, Manresa, and Pouliot (2023), showcasing its performance in a simulated single-graph environment based on a simple benchmark model.

The starting point is the adversarial formulation in which a discriminator separates observed objects (e.g., data) from analogous model-generated objects at parameter value θ . Originally introduced by Goodfellow et al. (2014) in the field of statistical learning,¹ the generative-adversarial framework has been adapted to structural

¹Goodfellow et al. (2014) show that, for a rich discriminator class, the concentrated criterion is a monotone transformation of Jensen–Shannon divergence; specifically, it equals $-\log 4 + 2\text{JS}(p_{\text{data}} \parallel p_g)$, where p_{data} and p_g are the data and generator densities.

econometric estimation by Kaji, Manresa, and Pouliot (2023). Their key idea is to use a structural model as the generator, simulated at candidate parameter values, resulting in a parameter estimate that makes observed data indistinguishable from synthetic ones. Simulations show that this approach overcomes multiple practical and computational limitations of typical structural estimators.

We specialize this adversarial framework to our single-graph setting of interest. To properly characterize cross-observation dependence, we represent the data through *ego objects*: for each node u in the network, this is an induced subgraph of radius k paired with covariates and equilibrium outcomes on its network neighborhood. The resulting ego objects overlap across nodes and are driven by a common equilibrium, so losses are dependent. The estimator we study is defined by the minimax problem:

$$\hat{\theta} \in \arg \min_{\theta \in \Theta} \sup_{D \in \mathcal{D}} \left\{ \mathbb{E}_{P_{\text{data}}}[\log D(S)] + \mathbb{E}_{P_{\theta}}[\log(1 - D(S))] \right\},$$

where Θ is the parameter space, \mathcal{D} is a discriminator class, S is an ego object, P_{data} denotes the observed distribution of local observations under the observed equilibrium, and P_{θ} denotes the corresponding model-implied distribution of local observations at parameter θ . In our setting, P_{data} and P_{θ} are the observed and simulated distributions of k ego objects. Intuitively, the estimator selects the parameter θ such that collections of subnetworks drawn from the observed graph are least distinguishable from the corresponding simulated subnetworks.

Our first theoretical contribution is an identification result. Having established well-posedness of equilibrium and measurability of local-observation distributions, we show that the population criterion is a constant plus twice the Jensen-Shannon divergence between the observed and simulated local-observation distributions, so the population minimizer equalizes them.² Identification follows by combining divergence faithfulness with a distributional separation argument: distinct structural parameters induce distinct focal-node-centered local distributions under the maintained assumptions. Our second theoretical contribution is about consistency. The logic of our proof is as follows. Under an ℓ_{∞} contraction with modulus $\rho \in (0, 1)$, shocks outside a radius- $(k + r)$ ball around focal node u affect outcomes inside the radius- k ball at rate ρ^{r+1} . This yields a truncated loss measurable with respect to shocks inside the

²This is closely related in spirit to minimum-distance and GMM formulations with large moment classes, but here the test-function search is embedded in the adversarial step rather than specified ex ante.

radius- $(k + r)$ ball. A dependence graph on focal nodes, defined by overlap of these balls, admits a bounded *coloring* (group partitioning) under bounded degree. Within each color class, truncation neighborhoods are disjoint, and truncated losses are conditionally independent. This construction yields a uniform law of large numbers for the empirical criterion and hence consistency.³

From a computational standpoint, a key contribution of the paper is to make adversarial (minimum-distance) structural estimation feasible when the data consist of a single large network equilibrium. The estimator can be implemented with three modular primitives: (i) a stable equilibrium solver based on Picard iteration under contraction, (ii) an ego-neighborhood extraction routine that turns the full network outcome into many local observations centered at focal nodes, and (iii) a flexible, permutation-invariant discriminator (for example, a message-passing architecture) which operates directly on these local graph objects. Training then proceeds by alternating updates: for fixed structural parameters, the discriminator is trained to separate observed and simulated ego neighborhoods; for fixed discriminator parameters, the structural parameters are updated by gradient methods using automatic differentiation through the unrolled equilibrium iteration. We emphasize practical stabilization devices (including input noise, gradient clipping, and low-overlap focal-node sampling) that mitigate overfitting and weak gradients on finite graphs. Taken together, these choices transform what would otherwise be an intractable likelihood problem on a massive dependent outcome vector into a scalable simulation-based routine whose per-iteration cost is driven mainly by sparse local computations on the graph and a modest number of equilibrium iterations. Our computational routine, packaged in Python, is publicly available through a dedicated repository.

The simulation results indicate that the adversarial estimator performs well in the benchmark linear-in-means environment and that the proposed diagnostics are informative in practice. In the representative run based on a synthetic graph with 249,813 nodes and 688,630 undirected edges, the final estimates are $\hat{\beta} = 0.4096$ and $\hat{\gamma} = 1.5046$, implying absolute errors of 0.0096 and 0.0046 relative to the true parameter values $(\beta_0, \gamma_0) = (0.4, 1.5)$. The parameter trajectories in Figure 1 converge from non-true initial values to a stable neighborhood of the truth, while the discriminator

³As in Kaji, Manresa, and Pouliot (2023), we focus on identification and consistency and advocate bootstrap-based inference in applications. A natural next step is to adapt dependence-robust network inference tools such as Kojevnikov, Marmer, and Song (2021) to this adversarial setting.

and structural-parameter losses in Figure 2 stabilize near their theoretical benchmark values $2\log 2$ and $\log 2$, respectively. Consistent with this, the final discriminator score distributions for observed and simulated ego objects overlap closely around $1/2$ in Figure 3, indicating that the discriminator can no longer reliably distinguish the two local-observation laws. Finally, the tail statistics in Table 3 show limited variability over the last 500 structural updates, with means close to the data-generating values, which further supports the conclusion that the training dynamics are stable and that the estimator achieves accurate finite-sample recovery in this simulation design.

This paper is situated in the growing literature on the econometrics of networks, originally motivated by peer effects problems. Most econometric identification and estimation results in this literature are developed in linear or linearized frameworks, including linear social-interaction and spatial autoregressive models (Bramoullé, Djebbari, and Fortin 2009; Lee, Liu, and Lin 2010; Lin and Lee 2010; Liu and Lee 2010; Liu 2014; Blume et al. 2015; Liu and Saraiva 2019; Liu 2020; Drukker, Egger, and Prucha 2023). These contributions are foundational, but their tractability typically relies on linear structure that may understate nonlinear propagation, strategic complementarities, and distributional asymmetries in equilibrium responses. Models with richer strategic and endogenous-network features point to the importance of nonlinear mechanisms (Kranton, D’Amours, and Bramoullé 2014; Johnsson and Moon 2021; Arduini, Patacchini, and Rainone 2015; de Paula, Rasul, and Sousa 2023), yet they generally do not provide an estimation-and-asymptotics framework targeted to a single observed graph with overlapping local objects. Our contribution is complementary: we keep the network observed and fixed, allow nonlinear structural equilibrium behavior, and develop theory and computation for the single-graph case using adversarial minimum-distance methods in the spirit of Kaji, Manresa, and Pouliot (2023).

This paper is also related to a growing literature that adapts neural networks for econometric estimation and inference. On the econometrics side, deep networks have been used as flexible sieve estimators with valid inferential guarantees (Farrell, Liang, and Misra 2021), to characterize individual endogeneity in structural models (Farrell, Liang, and Misra 2025), to handle endogeneity in counterfactual prediction (Hartford et al. 2017), and to implement moment-based estimation through adversarial and deep-GMM objectives (Lewis and Syrgkanis 2018; Bennett, Kallus, and Schnabel 2019; Kaji, Manresa, and Pouliot 2023). On networked data, recent work has begun

to adapt graph neural architectures to causal problems with interference and network confounding (Leung and Loupos 2025). Our contribution is complementary: we develop identification and consistency theory for adversarial structural estimation in a single-graph environment, using message-passing representations motivated by the graph-learning literature (Scarselli et al. 2009; Gilmer et al. 2017; Kipf and Welling 2017; Hamilton, Ying, and Leskovec 2017; Velickovic et al. 2018; Xu et al. 2019).

The remainder of the paper is organized as follows. Section 2 characterizes our single-graph environment, the associated statistical setup, and establishes identification. Section 3 illustrates our consistency result, and the assumptions that underpin it. Section 4 describes the estimation algorithm, highlighting its computational aspects. Section 5 assesses our estimator via simulations of a simple linear-in-means model. Lastly, Section 6 concludes. Three appendices supplement this paper: Appendix A collects mathematical proofs, Appendix B provides broader methodological and interpretive discussion, and Appendix C verifies the technical properties of the model used in the simulation.

2 Model and Identification

This section proceeds in four stages. Section 2.1 introduces the single-graph structural environment, local observations, and the minimum-distance criterion. Section 2.2 states the identifying assumptions. Section 2.3 presents representative economic models that satisfy these conditions. Section 2.4 establishes population identification; consistency is deferred to Section 3.

2.1 Statistical model

We observe a single realization (G_n, X_n, Y_n) , where $G_n = (V_n, E_n)$ is a simple undirected graph on n vertices, $X_n = (X_{n,1}, \dots, X_{n,n}) \in \mathcal{X}^n$ is a vector of node covariates, and $Y_n = (Y_{n,1}, \dots, Y_{n,n}) \in \mathbb{R}^n$ is an equilibrium outcome vector generated at a true parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^p$. We take \mathcal{X} to be a Borel subset of \mathbb{R}^{d_x} equipped with its Borel σ -algebra. Let $\mathcal{N}(\theta_0) \subseteq \mathbb{R}^p$ denote an open set containing Θ , used to state contraction and smoothness conditions. Throughout Section 2 we condition on the realized (G_n, X_n) and treat them as fixed.

For a vector v in a Euclidean space, $\|v\|$ denotes the Euclidean norm; in particular,

$\mathbb{S}^{p-1} := \{v \in \mathbb{R}^p : \|v\| = 1\}$. For $y \in \mathbb{R}^n$, $\|y\|_\infty := \max_{1 \leq i \leq n} |y_i|$, and for a subset $B \subseteq V_n$, $\|y\|_{\infty, B} := \max_{j \in B} |y_j|$. For a matrix $A \in \mathbb{R}^{n \times n}$, $\|A\|_\infty := \max_{1 \leq i \leq n} \sum_{j=1}^n |A_{ij}|$ is the induced operator norm. For $u \in V_n$ and integer $r \geq 0$, define $B_{n,r}(u) := \{v \in V_n : d_n(u, v) \leq r\}$, where d_n is the shortest path distance on G_n . Write $G_n[B_{n,r}(u)]$ for the induced subgraph and $\nu_n(r) := \max_{u \in V_n} |B_{n,r}(u)|$ for the maximal ball size at radius r .

Let $\varepsilon_n = (\varepsilon_{n,1}, \dots, \varepsilon_{n,n})$ denote the structural shocks. For each $\theta \in \Theta$, let Y_n^θ denote an outcome vector satisfying the fixed point equation

$$Y = T_{\theta,n}(Y, X_n, \varepsilon_n),$$

where $T_{\theta,n} : \mathbb{R}^n \times \mathcal{X}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the equilibrium map, with $T_{\theta,n,i}$ denoting its i th coordinate. The observed outcome equals $Y_n = Y_n^{\theta_0}$.

Fix an ego radius $k \geq 1$. The local observation at parameter θ for focal node $u \in V_n$ is the distance- k ego neighborhood centered at u :

$$\tilde{S}_{n,k}^\theta(u) := \left(G_n[B_{n,k}(u)], (X_{n,j}, Y_{n,j}^\theta)_{j \in B_{n,k}(u)}, u \right).$$

We represent $\tilde{S}_{n,k}^\theta(u)$ in a label-invariant way: permutations of non-focal nodes that preserve the focal-node-centered structure do not change the observation. The resulting label-invariant representation is denoted $S_{n,k}^\theta(u)$, and to define $P_{n,k}^\theta$ on a fixed measurable space we construct the corresponding label-invariant measurable representation space. Let \mathcal{G}_k be the finite set of focal-node-centered graph types with radius at most k and maximum degree at most Δ (from (A1)). For $g \in \mathcal{G}_k$, write $m(g) := |V(g)|$ and let $\text{Aut}_\circ(g)$ denote the finite within-neighborhood relabeling symmetry group (focal-node-fixed permutations) of g . Define the attribute fiber $\mathcal{A}_g := \mathcal{X}^{m(g)} \times \mathbb{R}^{m(g)}$ equipped with the product σ -algebra, and let $\text{Aut}_\circ(g)$ act on \mathcal{A}_g by permuting coordinates. Let $\pi_g : \mathcal{A}_g \rightarrow \mathcal{A}_g / \text{Aut}_\circ(g)$ be the orbit map and equip the quotient with the quotient σ -algebra

$$\mathcal{B}(\mathcal{A}_g / \text{Aut}_\circ(g)) := \{B \subseteq \mathcal{A}_g / \text{Aut}_\circ(g) : \pi_g^{-1}(B) \in \mathcal{B}(\mathcal{A}_g)\}.$$

Define the ego-object space

$$\mathcal{S}_k := \bigsqcup_{g \in \mathcal{G}_k} (\{g\} \times \mathcal{A}_g / \text{Aut}_\circ(g)),$$

equipped with the disjoint-union σ -algebra \mathcal{F}_k . For each focal node $u \in V_n$, let $g(u) \in \mathcal{G}_k$ denote the focal-node-centered graph type of $G_n[B_{n,k}(u)]$. Choose any focal-node-preserving isomorphism $\iota_u : g(u) \rightarrow G_n[B_{n,k}(u)]$. Pull back covariates and outcomes along ι_u to obtain an element $a_{n,k}^\theta(u) \in \mathcal{A}_{g(u)}$, and define

$$S_{n,k}^\theta(u) := (g(u), \pi_{g(u)}(a_{n,k}^\theta(u))) \in \mathcal{S}_k.$$

This definition does not depend on the choice of ι_u because different choices differ by an element of $\text{Aut}_\circ(g(u))$. For notational convenience, we write $(\mathcal{S}_{n,k}, \mathcal{F}_{n,k}) := (\mathcal{S}_k, \mathcal{F}_k)$.

The observable (graph, covariate) component of the ego object lives in the corresponding label-invariant representation space. For $g \in \mathcal{G}_k$, let $\text{Aut}_\circ(g)$ act on $\mathcal{X}^{m(g)}$ by coordinate permutation and let $\pi_{X,g} : \mathcal{X}^{m(g)} \rightarrow \mathcal{X}^{m(g)} / \text{Aut}_\circ(g)$ be the orbit map equipped with the quotient σ -algebra. Define the signature space

$$\mathcal{W}_k := \bigsqcup_{g \in \mathcal{G}_k} (\{g\} \times \mathcal{X}^{m(g)} / \text{Aut}_\circ(g))$$

with its disjoint-union σ -algebra, and define the measurable signature map $\Omega : \mathcal{S}_k \rightarrow \mathcal{W}_k$ by dropping the outcome coordinates. For each focal node u , write $\omega_u := \Omega(S_{n,k}^\theta(u))$, which is deterministic conditional on (G_n, X_n) and does not depend on θ .

Let $U_n \sim \text{Unif}(V_n)$ be a uniform random draw of a focal node, independent of ε_n . The k distribution of local observations at parameter θ is the probability measure on $(\mathcal{S}_{n,k}, \mathcal{F}_{n,k})$ defined by

$$P_{n,k}^\theta(\cdot) := \mathbb{P}_\theta(S_{n,k}^\theta(U_n) \in \cdot \mid G_n, X_n) = \frac{1}{n} \sum_{u \in V_n} \mathbb{P}_\theta(S_{n,k}^\theta(u) \in \cdot \mid G_n, X_n).$$

For each fixed focal node $u \in V_n$, define the unit-specific ego object law $P_{n,k}^{\theta,u} := \mathcal{L}(S_{n,k}^\theta(u) \mid G_n, X_n)$, the conditional distribution of the ego object at focal node u .

The probability measures \mathbb{P}_θ and expectations \mathbb{E}_θ are taken with respect to the randomness in the shock vector driving Y_n^θ and the auxiliary focal-node draw U_n ,

conditional on (G_n, X_n) . When both observed and simulated objects appear in the same expression, they are constructed using independent copies of these variables. The observed distribution of local observations is $P_{n,k}^{\theta_0}$. For a measurable adaptive test function $D : \mathcal{S}_{n,k} \rightarrow [0, 1]$, the population minimum-distance criterion is

$$V_{n,k}(\theta, D) := \mathbb{E}_{\theta_0}[\log D(S_{n,k}^{\theta_0}(U_n)) \mid G_n, X_n] + \mathbb{E}_{\theta}[\log(1 - D(S_{n,k}^{\theta}(U_n))) \mid G_n, X_n],$$

where the first expectation is over the observed equilibrium's shocks and U_n , and the second is over an independent simulation draw at θ and an independent uniform focal-node draw. The concentrated (profiled) minimum-distance criterion is

$$V_{n,k}^*(\theta) := \sup_{D: \mathcal{S}_{n,k} \rightarrow [0,1], \text{ meas.}} V_{n,k}(\theta, D),$$

with the convention $0 \cdot \log 0 := 0$.

2.2 Identification assumptions

We collect the assumptions required for the identification proof. The assumptions are grouped into three blocks: graph architecture and locality (A), behavioral-map and shock regularity (B), and differentiability plus relevance conditions that deliver parameter distinguishability (D).

Architecture and locality block (A). **(A1) Bounded degree.** Writing $\deg_n(u) := |\{v \in V_n : (u, v) \in E_n\}|$, there exists $\Delta < \infty$ such that $\max_{u \in V_n} \deg_n(u) \leq \Delta$ for all n .

This implies $|B_{n,k}(u)| \leq \nu_n(k) = O(\Delta^k) < \infty$ for each fixed k .⁴

(A2) Locality. For each $i \in V_n$, $T_{\theta,n,i}(y, x, \varepsilon)$ depends on (y, x, ε) only through coordinates indexed by $B_{n,1}(i)$.

This restriction implies that the i th coordinate of the equilibrium map depends only on the outcomes, covariates, and shocks in the radius one neighborhood of i . Combined with contraction, it yields decay of influence with graph distance, quantified in lemma 7.⁵

⁴For fixed k , the set of focal-node-centered k -ball graph label-invariant structural types \mathcal{G}_k is finite and each fiber $\mathcal{A}_g = \mathcal{X}^{m(g)} \times \mathbb{R}^{m(g)}$ has uniformly bounded dimension; this finiteness underlies the construction of the ego-object space $(\mathcal{S}_k, \mathcal{F}_k)$ in section 2.1.

⁵The results extend to interaction radius $r_0 \geq 1$ by replacing $B_{n,1}$ with B_{n,r_0} throughout.

Behavioral-map and shock block (B). **(B1) ℓ_∞ contraction.** There exists $\rho \in (0, 1)$ such that

$$\|T_{\theta,n}(y, x, \varepsilon) - T_{\theta,n}(y', x, \varepsilon)\|_\infty \leq \rho \|y - y'\|_\infty,$$

uniformly for $\theta \in \mathcal{N}(\theta_0)$, all $y, y' \in \mathbb{R}^n$, and all $(x, \varepsilon) \in \mathcal{X}^n \times \mathbb{R}^n$.⁶

This condition implies that, for each fixed $(n, \theta, x, \varepsilon)$, the map $y \mapsto T_{\theta,n}(y, x, \varepsilon)$ is a contraction in $\|\cdot\|_\infty$ with modulus ρ . It yields uniqueness of the equilibrium and geometric convergence of Picard iteration, and it is the main quantitative input for the influence decay and truncation bounds used below.⁷

(B2) Additive shocks.

$$T_{\theta,n}(y, x, \varepsilon) = h_{\theta,n}(y, x) + \varepsilon,$$

where $h_{\theta,n} : \mathbb{R}^n \times \mathcal{X}^n \rightarrow \mathbb{R}^n$ is measurable in all arguments and ρ -Lipschitz in y with respect to $\|\cdot\|_\infty$.

Additivity separates the deterministic interaction $h_{\theta,n}$ from the stochastic shocks and yields the contraction bound in lemma 1 by cancellation of shocks in differences. It is also used in the local approximation with influence decay of Section 3, where equilibria under two shock vectors are compared coordinatewise.

(B3) Exogeneity and smooth independent and identically distributed shocks. Conditional on (G_n, X_n) , the shocks $(\varepsilon_{n,i})_{i \in V_n}$ are independent and identically distributed (i.i.d.) and independent of (G_n, X_n) , with Lebesgue density f_ε on \mathbb{R} satisfying: (i) $f_\varepsilon \in C^1(\mathbb{R})$ and $f_\varepsilon(z) > 0$ for all $z \in \mathbb{R}$; (ii) $\mathbb{E}|\varepsilon_{n,1}| < \infty$; and (iii) $\int_{\mathbb{R}} |f'_\varepsilon(z)| dz < \infty$. We write the joint shock density as $f_{\varepsilon,n}(\varepsilon) := \prod_{i=1}^n f_\varepsilon(\varepsilon_i)$.

This assumption provides a strictly positive and differentiable shock density and conditional independence across vertices. These properties are used in Lemma 1⁸ to obtain a change of variables expression for the equilibrium density and, by marginalization, for the densities of finite outcome subvectors.⁸

⁶All subsequent applications evaluate this bound at the realized covariates $x = X_n$; the uniform-in- x statement is assumed for generality.

⁷In the linear in means model $Y = \beta WY + X\gamma + \varepsilon$ with row-stochastic W , the condition reduces to $|\beta| < 1$.

⁸The positivity of f_ε implies that the score function $\psi := f'_\varepsilon/f_\varepsilon$ is well defined.

Differentiability and relevance block (D). **(D1) Smoothness in states and parameters.** For each n and coordinate i , conditional on x , the map $(y, \theta) \mapsto h_{\theta, n, i}(y, x)$ is continuously differentiable in y and θ on $\mathbb{R}^n \times \mathcal{N}(\theta_0)$. Writing the state Jacobian as $A_{\theta, n}(y, x) := \partial_y h_{\theta, n}(y, x) \in \mathbb{R}^{n \times n}$, we assume: (i) $A_{\theta, n, ij}(y, x) = 0$ whenever $j \notin B_{n, 1}(i)$, and $\partial_\theta h_{\theta, n, i}(y, x)$ depends on (y, x) only through coordinates in $B_{n, 1}(i)$ (locality of derivatives); (ii) $\|A_{\theta, n}(y, X_n)\|_\infty \leq \rho < 1$ for all $y \in \mathbb{R}^n$ and all $\theta \in \mathcal{N}(\theta_0)$ (uniform Jacobian contraction); and (iii) the mixed derivative $\partial_\theta A_{\theta, n}(y, x)$ exists, is continuous in (y, θ) , inherits the same locality, and is uniformly bounded by $\|\partial_\theta A_{\theta, n}(y, x)\|_\infty \leq M_A$ (mixed derivative regularity).

Assumption (D1), combined with (B1) and (B2), implies that the map $F_{\theta, n}(y) := y - h_{\theta, n}(y, X_n)$ is a global C^1 diffeomorphism in Lemma 1'. This representation yields the density formula in Lemma 1" and justifies differentiation of the equilibrium with respect to θ through the implicit function formulas.

(D2) Distributional relevance. There exist constants $c_0 > 0$, $c_g > 0$, and $t_0 > 0$ such that for every direction $\delta \in \mathbb{S}^{p-1}$, there exists a set $U_\delta \subseteq V_n$ with $|U_\delta| \geq c_0 n$ such that for each $u \in U_\delta$, writing $B := B_{n, k}(u)$, the density $f_{B|\theta, u}$ from corollary 1 admits a uniform first order expansion along δ at θ_0 as an $L^1(\mathbb{R}^{|B|})$ -valued map. That is, there exists $g_{u, \delta} \in L^1(\mathbb{R}^{|B|})$ with $\|g_{u, \delta}\|_{L^1} \geq c_g$ such that for all t with $|t| \leq t_0$,

$$\|f_{B|\theta_0+t\delta, u}(\cdot) - f_{B|\theta_0, u}(\cdot) - tg_{u, \delta}(\cdot)\|_{L^1(\mathbb{R}^{|B|})} \leq \frac{c_g}{2}|t|.$$

Assumption (D2) is a relevance condition stated in terms of the derived local densities. It implies a uniform local separation in total variation (TV) along each direction.⁹ Hence for $u \in U_\delta$ and $0 < |t| \leq t_0$,

$$\|L_u^{\theta_0+t\delta} - L_u^{\theta_0}\|_{\text{TV}} = \frac{1}{2}\|f_{B|\theta_0+t\delta, u} - f_{B|\theta_0, u}\|_{L^1} \geq \frac{c_g}{4}|t|.$$

This condition rules out cancellations under marginalization to the k ego subvector that can eliminate first order variation of the local density.

(D3) Distinguishable signatures. Fix the ego radius k and write $\omega_u :=$

⁹If P and Q are absolutely continuous with respect to a common dominating measure λ with densities p and q , then $\|P - Q\|_{\text{TV}} = \frac{1}{2}\|p - q\|_{L^1(\lambda)}$; see Axler (2020, Theorem 9.10). Applying this identity with $P = L_u^{\theta_0+t\delta}$ and $Q = L_u^{\theta_0}$ gives $\|L_u^{\theta_0+t\delta} - L_u^{\theta_0}\|_{\text{TV}} = \frac{1}{2}\|f_{B|\theta_0+t\delta, u} - f_{B|\theta_0, u}\|_{L^1} \geq \frac{c_g}{4}|t|$ for $u \in U_\delta$ and $0 < |t| \leq t_0$.

$\Omega(S_{n,k}^\theta(u)) \in \mathcal{W}_k$ for the observable signature of focal node u (defined in section 2.1), which consists of the focal-node-centered k -ball graph type together with the covariate configuration modulo focal-node-preserving relabeling. Assume that the signatures are distinct across focal nodes: for all $u \neq v$, $\omega_u \neq \omega_v$. Equivalently, for any distinct focal nodes $u \neq v$, if $G_n[B_{n,k}(u)]$ and $G_n[B_{n,k}(v)]$ are isomorphic via a focal-node-preserving relabeling, then the covariate configurations differ under every such relabeling.

This condition makes the map $u \mapsto \omega_u$ injective, so that each unit-specific outcome law L_u^θ can be recovered from the mixture distribution of local observations $F_{n,k}^\theta$ by conditioning on ω_u (formalized in lemma 5 below). It is used in theorem 1 to translate focal node-level variation of the outcome density into variation of the mixture distribution of local observations.¹⁰

2.3 Economic examples

The framework is intentionally broad. The following examples illustrate how standard network-econometric models map into the primitives and assumptions stated above.

Example 1 (Linear in means). Let

$$Y_n = \beta W_n Y_n + X_n \gamma + \varepsilon_n,$$

where W_n is row-stochastic and sparse along graph edges. Then $h_{\theta,n}(y, x) = \beta W_n y + x \gamma$ with $\theta = (\beta, \gamma)$. Assumption (A2) holds when $W_{n,ij} = 0$ outside the interaction neighborhood. Assumption (B1) holds under $|\beta| < 1$ (more generally $|\beta| \|W_n\|_\infty < 1$). Assumptions (B2) and (D1) are immediate.

Example 2 (Nonlinear best responses with strategic complementarities).

For each $i \in V_n$, let

$$Y_{n,i} = \sigma\left(\alpha + X'_{n,i} \gamma + \beta \bar{Y}_{n,i}\right) + \varepsilon_{n,i}, \quad \bar{Y}_{n,i} := \frac{1}{|B_{n,1}(i) \setminus \{i\}|} \sum_{j \in B_{n,1}(i) \setminus \{i\}} Y_{n,j},$$

¹⁰Under bounded degree and fixed k , if covariates take values in a finite set, then \mathcal{W}_k is finite, so injectivity fails for any n exceeding the number of possible signatures.

with a C^1 response function σ , and with the convention $\bar{Y}_{n,i} = 0$ when $B_{n,1}(i) \setminus \{i\}$ is empty. If $\sup_t |\sigma'(t)| \leq L_\sigma$ and $|\beta|L_\sigma < 1$, then the equilibrium map is a contraction in $\|\cdot\|_\infty$, so (B1) holds. Local averaging implies (A2), additivity gives (B2), and smoothness of σ yields (D1).

Example 3 (Saturating nonlinear peer aggregation). Let

$$Y_{n,i} = \phi_\theta \left(X_{n,i}, \sum_{j \in B_{n,1}(i) \setminus \{i\}} a_{ij} g(Y_{n,j}) \right) + \varepsilon_{n,i},$$

where g and ϕ_θ are C^1 and a_{ij} are bounded neighborhood weights. If g is Lipschitz with constant L_g and $\sup_{x,z,\theta} \|\partial_z \phi_\theta(x,z)\| \leq L_\phi$ with $L_\phi L_g \sup_i \sum_j |a_{ij}| < 1$, then (B1) holds. The model is local by construction (A2), additive in shocks (B2), and differentiable in states and parameters (D1).

Section 5 specializes estimation and finite-sample diagnostics to Example 1; the identification argument in this section is stated for the general setup.

2.4 Main identification result

This subsection establishes that the population minimum-distance criterion uniquely identifies the true parameter. Under assumptions (A1), (A2), (B1) through (B3), and (D1) through (D3), we show that the parameter vector θ_0 is the unique local minimizer of the concentrated population minimum-distance criterion. The result is a population identification theorem: it characterizes the target of the criterion, not the rate at which a sample estimator converges. Consistency, which additionally requires growing graph asymptotics, local approximation with influence decay, and adaptive test function complexity control, is deferred to Section 3.

The argument proceeds in five steps, each producing an object or property used by the next. First, contraction of the equilibrium map yields a unique measurable equilibrium for every parameter value (lemma 1). Second, bounded degree ensures that the induced map $\theta \mapsto P_{n,k}^\theta$ is well defined (lemma 2). Third, the concentrated minimum-distance criterion equals, up to a constant, the Jensen–Shannon divergence between the observed and simulated distributions of local observations (lemma 3). Fourth, faithfulness of the Jensen–Shannon divergence converts vanishing divergence into equality of distributions of local observations (lemma 4). Fifth, we show that

distinct parameters produce distinct distributions of local observations, completing identification (theorem 1).

The first four steps establish, under our equilibrium and locality assumptions, that the concentrated population minimum-distance criterion is a constant plus twice the Jensen–Shannon divergence between the observed and simulated k distributions of local observations. The final step shows that equality of k distributions of local observations implies $\theta = \theta_0$ under additional smoothness, distributional relevance, and observable heterogeneity conditions. The proof constructs a global diffeomorphism from shocks to equilibrium outcomes, uses it to derive a strictly positive equilibrium density and its finite-dimensional marginals, and then shows that these derived local densities vary with θ at a positive fraction of focal nodes. The contraction rate ρ enters quantitatively because it controls how much information about θ reaches a focal node through paths longer than k .

We now state and prove the lemmas that build to the main identification theorem. The argument begins by establishing the existence and regularity of the equilibrium, then links the minimum-distance criterion to the Jensen–Shannon divergence, and finally shows that distinct parameters yield distinct distributions of local observations.

Lemma 1 (Equilibrium). *Under (B1) and (B2), for every $\theta \in \mathcal{N}(\theta_0)$ and every realization of $(G_n, X_n, \varepsilon_n)$: (a) the equation $Y = h_{\theta,n}(Y, X_n) + \varepsilon_n$ has a unique solution $Y_n^\theta \in \mathbb{R}^n$; (b) the Picard iterates $Y^{(t+1)} := h_{\theta,n}(Y^{(t)}, X_n) + \varepsilon_n$ satisfy $\|Y^{(t)} - Y_n^\theta\|_\infty \leq \rho^t \|Y^{(0)} - Y_n^\theta\|_\infty$; and (c) the map $\varepsilon_n \mapsto Y_n^\theta$ is Borel measurable conditional on (G_n, X_n) .*

Proof. See Appendix A.

Up to this point, the argument uses only contraction to guarantee that the equilibrium is well defined and measurable. For identification, however, we must also control how the law of equilibrium outcomes varies with θ . Rather than assuming that the equilibrium admits a regular density, we derive this property from primitives. The key idea is to view the shock vector ε_n as the primitive random object and to treat the equilibrium equation $Y_n^\theta = h_{\theta,n}(Y_n^\theta, X_n) + \varepsilon_n$ as an invertible change of variables between ε_n and Y_n^θ .

Lemma 1' (Global shock to equilibrium diffeomorphism). *Fix n and condition on (G_n, X_n) . Under (B1), (B2), and (D1), for each $\theta \in \mathcal{N}(\theta_0)$ the map $F_{\theta,n}(y) :=$*

$y - h_{\theta,n}(y, X_n)$ is a C^1 diffeomorphism of \mathbb{R}^n onto itself. In particular, $Y_n^\theta = F_{\theta,n}^{-1}(\varepsilon_n)$, and the equilibrium depends smoothly on both ε_n and θ :

$$\partial_\varepsilon Y_n^\theta = (I_n - A_{\theta,n}(Y_n^\theta, X_n))^{-1}, \quad \partial_\theta Y_n^\theta = (I_n - A_{\theta,n}(Y_n^\theta, X_n))^{-1} \partial_\theta h_{\theta,n}(Y_n^\theta, X_n). \quad (2.1)$$

Proof. See Appendix A.

With the diffeomorphism in hand, we can derive the equilibrium density from the shock density by the change of variables formula.

Lemma 1'' (Equilibrium density). *Fix n and condition on (G_n, X_n) . Under (B1), (B2), (B3), and (D1), the equilibrium outcome vector Y_n^θ admits a strictly positive Lebesgue density on \mathbb{R}^n given by*

$$f_{Y|\theta,n}(y) = \left(\prod_{i=1}^n f_\varepsilon(y_i - h_{\theta,n,i}(y, X_n)) \right) |\det(I_n - A_{\theta,n}(y, X_n))|. \quad (2.2)$$

Moreover, $f_{Y|\theta,n}$ is continuous in (y, θ) and, under the mixed derivative bound in (D1)(iii), continuously differentiable in θ for each fixed y .

Proof. See Appendix A.

The full density (2.2) immediately yields regularity of finite-dimensional marginals by integration.

Corollary 1 (Local densities). *Fix n and condition on (G_n, X_n) . Let $u \in V_n$ and $B := B_{n,k}(u)$. Under (B1), (B2), (B3), and (D1), the marginal law $L_u^\theta := \mathcal{L}((Y_{n,j}^\theta)_{j \in B} \mid G_n, X_n)$ admits a Lebesgue density on $\mathbb{R}^{|B|}$ given by*

$$f_{B|\theta,u}(y_B) = \int_{\mathbb{R}^{n-|B|}} f_{Y|\theta,n}(y_B, y_{-B}) dy_{-B}. \quad (2.3)$$

In particular, each unit-specific ego object law $P_{n,k}^{\theta,u}$ is absolutely continuous on its outcome coordinates, and the k distribution of local observations $P_{n,k}^\theta = \frac{1}{n} \sum_u P_{n,k}^{\theta,u}$ is absolutely continuous with respect to counting measure on observable signatures times Lebesgue measure on outcomes.

Having established uniqueness of the equilibrium and regularity of the ego object space, we can now confirm that the parameter to distribution of local observations map is well defined.

Lemma 2 (Distribution of local observations). *Under (A1), (A2), (B1), and (B2), for each $\theta \in \mathcal{N}(\theta_0)$ the k distribution of local observations $P_{n,k}^\theta$ is a well defined probability measure on $(\mathcal{S}_{n,k}, \mathcal{F}_{n,k})$, and the map $\theta \mapsto P_{n,k}^\theta$ is well defined on $\mathcal{N}(\theta_0)$.*

Proof. See Appendix A.

We now connect the concentrated minimum-distance criterion to a statistical divergence. Lemma 3 shows that the concentrated minimum-distance criterion equals a constant plus twice the Jensen–Shannon divergence between the observed and simulated distributions of local observations.

Lemma 3 (Concentrated criterion and Jensen–Shannon divergence). *Let $(\mathcal{S}, \mathcal{F})$ be a measurable space and let P and Q be probability measures on it. For any measurable $D : \mathcal{S} \rightarrow [0, 1]$, define*

$$V(P, Q; D) := \mathbb{E}_P[\log D(S)] + \mathbb{E}_Q[\log(1 - D(S))],$$

where we adopt the conventions $0 \cdot \log 0 := 0$ and $0 \cdot \log(1 - 0) := 0$ (so $V(P, Q; D) \in [-\infty, 0]$ is well defined as an extended real number). Define the concentrated value

$$V^*(P, Q) := \sup_{D: \mathcal{S} \rightarrow [0,1]} V(P, Q; D).$$

Then: (a) the supremum is attained by the measurable function

$$D^*(s) = \frac{dP}{d(P+Q)}(s) \quad (P+Q)\text{-a.e. } s.$$

and (b) writing $M := \frac{1}{2}(P+Q)$, the concentrated value satisfies

$$V^*(P, Q) = -\log 4 + 2 \text{JS}(P \| Q),$$

where $\text{JS}(P \| Q) := \frac{1}{2} \text{KL}(P \| M) + \frac{1}{2} \text{KL}(Q \| M)$ and KL denotes the Kullback–Leibler divergence.

Proof. See Appendix A.

Setting $P = P_{n,k}^{\theta_0}$ and $Q = P_{n,k}^\theta$, we obtain $V_{n,k}^*(\theta) = -\log 4 + 2 \text{JS}(P_{n,k}^{\theta_0} \| P_{n,k}^\theta)$, so that $V_{n,k}^*(\theta) \geq -\log 4$ for all θ , with the lower bound attained when and only when the Jensen–Shannon divergence vanishes. The Jensen–Shannon divergence is faithful in the sense that it vanishes if and only if its arguments coincide:

Lemma 4 (Faithfulness of Jensen–Shannon divergence). *For any probability measures P and Q on a common measurable space, $\text{JS}(P \parallel Q) = 0$ if and only if $P = Q$.*

Proof. See Appendix A.

Combining lemmas 3 and 4, for any $\theta \in \mathcal{N}(\theta_0)$,

$$V_{n,k}^*(\theta) = -\log 4 \iff \text{JS}(P_{n,k}^{\theta_0} \parallel P_{n,k}^\theta) = 0 \iff P_{n,k}^\theta = P_{n,k}^{\theta_0}.$$

The population minimum-distance criterion achieves its minimum if and only if the k distribution of local observations at θ matches the k distribution of local observations at θ_0 . It remains to show that this matching forces $\theta = \theta_0$.

Lemma 5 (Mixture components under distinguishable signatures). *Fix $k \geq 1$ and condition on (G_n, X_n) . For each focal node $u \in V_n$, let $\omega_u := \Omega(S_{n,k}^\theta(u)) \in \mathcal{W}_k$ be the observable signature, which is deterministic conditional on (G_n, X_n) and does not depend on θ . Then the k distribution of local observations is the uniform mixture of unit-specific laws,*

$$P_{n,k}^\theta = \frac{1}{n} \sum_{u \in V_n} P_{n,k}^{\theta,u}, \quad P_{n,k}^{\theta,u} := \mathcal{L}(S_{n,k}^\theta(u) \mid G_n, X_n),$$

and each $P_{n,k}^{\theta,u}$ is supported on $\Omega^{-1}(\{\omega_u\})$. If the signatures are distinct across focal nodes (assumption (D3)), then for every focal node u and every measurable set $A \in \mathcal{F}_{n,k}$,

$$P_{n,k}^\theta(A \cap \Omega^{-1}(\{\omega_u\})) = \frac{1}{n} P_{n,k}^{\theta,u}(A).$$

Consequently, $P_{n,k}^\theta = P_{n,k}^{\theta_0}$ implies $P_{n,k}^{\theta,u} = P_{n,k}^{\theta_0,u}$ for all $u \in V_n$.

Proof. See Appendix A.

Theorem 1 (Identification). *Under (A1), (A2), (B1), (B2), (B3), and (D1), suppose there exists an ego radius $k_0 \geq 1$ such that conditions (D2) and (D3) hold at radius k_0 . Let t_0 denote the constant from (D2) at radius k_0 . Then for all $k \geq k_0$, θ_0 is the unique local minimizer of the population minimum-distance criterion on $\{\theta \in \Theta : \|\theta - \theta_0\| < t_0\}$: for all $\theta \in \Theta$ with $\|\theta - \theta_0\| < t_0$,*

$$V_{n,k}^*(\theta) \geq V_{n,k}^*(\theta_0) = -\log 4,$$

with equality if and only if $\theta = \theta_0$.

Proof. See Appendix A.

The proof decomposes into two logically independent parts. The chain of lemmas 1 to 4 establishes, for any structural model satisfying (A1), (A2), (B1), and (B2), that $V_{n,k}^*(\theta) = -\log 4$ if and only if $P_{n,k}^\theta = P_{n,k}^{\theta_0}$. This reduces local identification to showing that equality of k distributions of local observations forces $\theta = \theta_0$ in a neighborhood of θ_0 . Condition (D1) delivers well defined local densities. Condition (D2) provides a uniform local separation of at least one unit-specific marginal density along each parameter direction. Condition (D3) then lifts this separation from a unit-specific marginal to the mixture distribution of local observations. The consistency step, which requires growing graph asymptotics, local approximation with influence decay, and adaptive test function complexity control, is taken up in Section 3. Theorem 1 is stated for $k \geq k_0$ because conditions (D2) and (D3) are imposed at radius k_0 and need not hold at smaller radii.

3 Consistency

This section establishes that approximate minimizers of the empirical criterion converge to that target under growing-graph asymptotics with cross-observation dependence. The section is divided into four parts: Section 3.1 introduces the empirical criterion and proof strategy, section 3.2 states the consistency assumptions, section 3.3 develops the intermediate lemmas, and section 3.4 states the final consistency theorem.

3.1 Roadmap

In our setting, verifying consistency requires showing that the adversarial estimator converges to θ_0 as the network grows. The core difficulty is that the summands in the empirical criterion are dependent because ego neighborhoods centered at nearby nodes share vertices and are driven by common shocks. Thus, uniform laws of large numbers for (in)dependent observations do not apply directly. It is useful to develop here a roadmap of our asymptotic analysis.

Fix an ego radius $k \geq 1$. For each n , let Φ_n be an index set and let $\{D_\phi : \phi \in \Phi_n\}$ be a test-function sieve with $D_\phi : \mathcal{S}_{n,k} \rightarrow [0, 1]$. Let $\tilde{\varepsilon}_n$ be an independent copy

of the shock vector ε_n conditional on (G_n, X_n) , and let \tilde{Y}_n^θ denote the equilibrium outcome vector generated at parameter θ using shocks $\tilde{\varepsilon}_n$. Define the simulated ego object $\tilde{S}_{n,k}^\theta(u)$ to be the focal-node-preserving label-invariant representation obtained by replacing the outcome coordinates Y_n^θ with \tilde{Y}_n^θ in the labeled representative $\tilde{S}_{n,k}^\theta(u)$. For $(\theta, \phi) \in \Theta \times \Phi_n$, define the empirical criterion

$$\widehat{V}_{n,k}(\theta, \phi) := \frac{1}{n} \sum_{u \in V_n} \left\{ \log D_\phi(S_{n,k}^{\theta_0}(u)) + \log(1 - D_\phi(\tilde{S}_{n,k}^\theta(u))) \right\},$$

where we write the corresponding per-focal node loss as $\ell_{\theta,\phi}(u) := \log D_\phi(S_{n,k}^{\theta_0}(u)) + \log(1 - D_\phi(\tilde{S}_{n,k}^\theta(u)))$. For $r \geq 0$, let $\ell_{\theta,\phi}^{(u,r)}(u)$ denote the truncated counterpart constructed in lemma 8. Define its conditional expectation

$$V_{n,k}(\theta, \phi) := \mathbb{E}[\widehat{V}_{n,k}(\theta, \phi) \mid G_n, X_n].$$

The sieve criteria are $\widehat{V}_{n,k}^{\Phi_n}(\theta) := \sup_{\phi \in \Phi_n} \widehat{V}_{n,k}(\theta, \phi)$ and $V_{n,k}^{\Phi_n}(\theta) := \sup_{\phi \in \Phi_n} V_{n,k}(\theta, \phi)$. We write $\hat{\theta}_{n,k}$ for an approximate minimizer of $\widehat{V}_{n,k}^{\Phi_n}$ as specified in theorem 2. Throughout Section 3, probabilities and expectations are taken with respect to the joint conditional law of

$$\left(\varepsilon_n, \tilde{\varepsilon}_n, (\varepsilon_n^{(u,\text{out})})_{u \in V_n}, (\tilde{\varepsilon}_n^{(u,\text{out})})_{u \in V_n} \right) \text{ given } (G_n, X_n),$$

where $\tilde{\varepsilon}_n$ is an independent copy of ε_n , and the collections $\{\varepsilon_n^{(u,\text{out})}\}_{u \in V_n}$ and $\{\tilde{\varepsilon}_n^{(u,\text{out})}\}_{u \in V_n}$ are independent conditional on (G_n, X_n) , each consisting of independent copies of the full shock vector with the same conditional law as ε_n . Write

$$\mathcal{E}_n := \sigma(\varepsilon_n, \tilde{\varepsilon}_n, (\varepsilon_n^{(u,\text{out})})_{u \in V_n}, (\tilde{\varepsilon}_n^{(u,\text{out})})_{u \in V_n})$$

for the σ -algebra generated by these shocks.

Our strategy has three components: truncation, blocking, and concentration. The truncation step, powered by the contraction and locality assumptions already exploited in Section 2, replaces each ego neighborhood outcome vector by a local approximation that depends only on shocks within a radius $(k + r)$ ball, incurring an error of order ρ^{r+1} . A blocking partition of the resulting dependency structure partitions the truncated summands into finitely many conditionally independent groups. Concentration bounds together with a covering argument then yield uniform con-

vergence of the concentrated empirical minimum-distance criterion to its population counterpart, and a separation argument yields consistency.

The argument proceeds in seven steps, grouped into three blocks. Lemmas 6 to 8 control the truncation approximation. Lemmas 9 to 11 establish uniform convergence via blocking, concentration, and a covering argument, with (C3') providing the effective Lipschitz control needed for the off-net extension. Theorem 2 then combines uniform convergence with the identification result of Section 2.

3.2 Assumptions

We now state conditions for argmin consistency of approximate minimizers of $\widehat{V}_{n,k}^{\Phi_n}$. Relative to Section 2, the additional ingredients are: control of overlap dependence across focal nodes, complexity control of the sieve Φ_n , and sequence-uniform population separation. Assumptions (A1), (A2), (B1), and (B2) carry over unchanged; we retain the same notation and numbering from section 2.2. The smooth positive-density part of (B3) is not needed for the stochastic part of the consistency argument; conditional shock independence and tail control enter through (C4). Fix a truncation sequence $(r_n)_{n \geq 1}$ with $r_n \uparrow \infty$, chosen to satisfy the tail and effective-sample conditions in theorem 2.

(C1) Compactness. The parameter space $\Theta \subset \mathbb{R}^p$ is compact and $\theta_0 \in \text{int}(\Theta)$.

Compactness ensures that Θ can be covered by finitely many balls of any fixed radius, which is used in the covering argument for uniform convergence.¹¹

(C2) Uniform bounded anchor. There exists $M_h < \infty$ such that $\|h_{\theta,n}(0, X_n)\|_\infty \leq M_h$ for all n and all $\theta \in \Theta$.

This bounds the value of the structural interaction at the zero outcome vector. It is a joint restriction on the parameter space and covariates.¹² Together with contraction, it controls the magnitude of equilibrium outcomes and the tail of the Picard iteration, ensuring that the local approximation with influence decay produces finite approximation errors.

(C3) Lipschitz dependence on parameters. There exists $L_\theta < \infty$ such that for all n , all $\theta, \theta' \in \Theta$, and all $y \in \mathbb{R}^n$,

¹¹The interior condition $\theta_0 \in \text{int}(\Theta)$ excludes boundary cases in the approximate argmin argument.

¹²In the linear in means model $h_{\theta,n}(y, x) = \beta W_n y + X_n \gamma$, this reduces to $\|X_n \gamma\|_\infty \leq M_h$. A sufficient condition is bounded covariates together with compact Θ .

$$\|h_{\theta,n}(y, X_n) - h_{\theta',n}(y, X_n)\|_\infty \leq L_\theta \|\theta - \theta'\|.$$

This deterministic smoothness condition implies a Lipschitz bound for the equilibrium in θ via lemma 6(b) and is a sufficient condition for the effective Lipschitz property (C3') used in the covering argument in lemma 10.¹³

(C3') Effective Lipschitz dependence on parameters. There exist events $\mathcal{A}_n \in \mathcal{E}_n$ with $\mathbb{P}(\mathcal{A}_n | G_n, X_n) \rightarrow 1$ and deterministic constants $L_n < \infty$ such that on \mathcal{A}_n , for all $\theta, \theta' \in \Theta$, all $\phi \in \Phi_n$, and all $u \in V_n$,

$$|\ell_{\theta,\phi}^{(u,r_n)}(u) - \ell_{\theta',\phi}^{(u,r_n)}(u)| \leq L_n \|\theta - \theta'\|,$$

where $\ell_{\theta,\phi}^{(u,r_n)}(u)$ is the truncated loss from lemma 8 at depth r_n . If (C3) holds, then (C3') holds with $\mathcal{A}_n = \Omega$ and $L_n = L_D L_\theta / (\eta(1 - \rho))$ by lemma 6(b) and (C6)(i).

(C4) Shock tail control. Conditional on (G_n, X_n) , the shocks $(\varepsilon_{n,i})_{i \in V_n}$ are i.i.d. and the simulated shocks $(\tilde{\varepsilon}_{n,i})_{i \in V_n}$ are an independent i.i.d. copy. For a truncation sequence $r_n \uparrow \infty$ (specified in the theorem),

$$\rho^{r_n} \|\varepsilon_n\|_\infty \xrightarrow{p} 0, \quad \rho^{r_n} \|\tilde{\varepsilon}_n\|_\infty \xrightarrow{p} 0,$$

and $\rho^{r_n} \mathbb{E}[\|\varepsilon_n\|_\infty | G_n, X_n] \rightarrow 0$.

This condition ensures that the product $\rho^{r_n} \|\varepsilon_n\|_\infty$ vanishes in probability.¹⁴

(C5) Ego object metric. We define d_S to be the product metric on $\mathcal{S}_{n,k}$ that is discrete on the focal-node-centered graph and covariate components and uses the ℓ_∞ norm on the outcome coordinates. In particular, for all n , all $u \in V_n$, and all outcome vectors $y, y' \in \mathbb{R}^n$,

$$d_S(S_{n,k}(u; y), S_{n,k}(u; y')) \leq \|y - y'\|_{\infty, B_{n,k}(u)},$$

where $S_{n,k}(u; y)$ denotes the ego object at focal node u with outcomes y and $\|y - y'\|_{\infty, B_{n,k}(u)} := \max_{j \in B_{n,k}(u)} |y_j - y'_j|$.

¹³A sufficient primitive condition is the existence of $M_\theta < \infty$ such that $\|\partial_\theta h_{\theta,n,i}(y, X_n)\| \leq M_\theta$ uniformly in (n, θ, y, i) , which implies (C3) with $L_\theta = M_\theta$ by the mean value theorem (Lebl 2025, Theorem 4.2.4). In models such as linear in means, $\partial_\beta h_{\theta,n}(y, X_n) = W_n y$ is unbounded over $y \in \mathbb{R}^n$, so (C3) fails and one uses the weaker high-probability condition (C3').

¹⁴A practical check is to pick r_n such that $\rho^{r_n} a_n \rightarrow 0$ for a deterministic sequence a_n with $\|\varepsilon_n\|_\infty = O_p(a_n)$. The additional requirement $\rho^{r_n} \mathbb{E}[\|\varepsilon_n\|_\infty | G_n, X_n] \rightarrow 0$ controls the population truncation bias in lemma 11.

To make the discrete-continuous product structure compatible with the quotient formalism of section 2.1, we define d_S as follows. For $s, s' \in \mathcal{S}_k$, set $d_S(s, s') := 1$ if $\Omega(s) \neq \Omega(s')$. If $\Omega(s) = \Omega(s')$, then s and s' lie in the same focal-node-centered graph fiber g . Choose representatives $(x, y), (x', y') \in \mathcal{A}_g$ such that $\pi_g(x, y) = s$ and $\pi_g(x', y') = s'$, and define

$$d_S(s, s') := \min\{\|y - \pi y'\|_\infty : \pi \in \text{Aut}_o(g), x = \pi x'\}.$$

This definition does not depend on the choice of representatives and coincides with the ℓ_∞ distance between outcome subvectors whenever the signature has trivial stabilizer.¹⁵

For a bounded measurable function $f : \mathcal{S}_{n,k} \rightarrow \mathbb{R}$, we write $\|f\|_\infty := \sup_{s \in \mathcal{S}_{n,k}} |f(s)|$.

(C6) Adaptive test function regularity. Adaptive test functions $D_\phi : \mathcal{S}_{n,k} \rightarrow [\eta, 1 - \eta]$ for a fixed clipping level $\eta \in (0, 1/2)$ satisfy: (i) for all $\phi \in \Phi_n$ and all $s, s' \in \mathcal{S}_{n,k}$,

$$|D_\phi(s) - D_\phi(s')| \leq L_D d_S(s, s');$$

and (ii) the function class $\mathcal{F}_n := \{s \mapsto \log D_\phi(s), s \mapsto \log(1 - D_\phi(s)) : \phi \in \Phi_n\}$ satisfies, for each fixed $\epsilon > 0$,

$$\log N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) = o(m_{\text{eff},n}),$$

where $N(\epsilon, \cdot, \|\cdot\|_\infty)$ denotes the covering number under $\|\cdot\|_\infty$ and $m_{\text{eff},n} := m_{\text{eff},n}(r_n)$ is the effective number of independent local observations defined in (3.3) after lemma 9.

The Lipschitz condition in (i) ensures that small perturbations of outcomes produce small changes in discriminator output, which is needed to pass from the truncation bound on outcomes to a truncation bound on the adversarial loss.¹⁶

(C7) Sieve approximation. The sieve approximation error

$$\alpha_n := \sup_{\theta \in \Theta} |V_{n,k}^*(\theta) - V_{n,k}^{\Phi_n}(\theta)|$$

¹⁵In all cases, taking $\pi = \text{id}$ yields the displayed inequality for fixed-focal node comparisons.

¹⁶Clipping to $[\eta, 1 - \eta]$ bounds $|\log D_\phi|$ and $|\log(1 - D_\phi)|$ by $|\log \eta|$, yielding bounded summands for concentration. The entropy condition in (ii) ensures that the test-function class is not too large relative to the effective number of independent local observations.

satisfies $\alpha_n \rightarrow 0$, where $V_{n,k}^*(\theta)$ is the infinite-capacity concentrated minimum-distance criterion. Let

$$V_{n,k}^{\Phi_n}(\theta) := \sup_{\phi \in \Phi_n} V_{n,k}(\theta, \phi)$$

denote the sieve restricted counterpart.

This requires that the test-function sieve Φ_n grows rich enough to approximate the optimal adaptive test function uniformly over θ in terms of the concentrated minimum-distance criterion.

Theorem 2 requires a population separation input that is uniform in n .¹⁷ We record the conclusion as a proposition and defer its proof to the appendix.

Proposition 1 (Uniform separation (C8)). *Fix an ego radius $k \geq 1$. Assume (A1), (A2), (B1)–(B3), (C1), and (D1). If the additional limit conditions (P1)–(P3) in Section A hold, then the following holds. For every $\epsilon > 0$ there exist $\delta(\epsilon) > 0$ and $N(\epsilon) \in \mathbb{N}$ such that for all $n \geq N(\epsilon)$,*

$$\inf_{\theta \in \Theta: \|\theta - \theta_0\| \geq \epsilon} (V_{n,k}^*(\theta) - V_{n,k}^*(\theta_0)) \geq \delta(\epsilon).$$

3.3 Intermediate results

This subsection provides the technical building blocks for consistency. Lemmas 6 to 8 control truncation bias; lemma 9 converts overlap dependence into a block structure and defines an effective sample size; lemmas 10 and 11 and corollary 2 then deliver uniform convergence of the concentrated criterion. We start with deterministic equilibrium properties that strengthen lemma 1. Under contraction and the bounded-anchor condition, the equilibrium is unique, uniformly bounded, and Lipschitz in the structural parameter. Following the convention of the lemma 1 proof, for fixed (X_n, ε) write $\Gamma_{\theta,n}(y) := h_{\theta,n}(y, X_n) + \varepsilon$ for the Picard map, reserving $T_{\theta,n}$ for the three-argument equilibrium map defined in Section 2.1.

Lemma 6 (Equilibrium boundedness and parameter sensitivity). *Under (B1), (B2), and (C2):*

¹⁷Because the graph G_n varies with n , $(V_{n,k}^*)_{n \geq 1}$ is a triangular array, and pointwise identification at fixed n (theorem 1) does not itself provide a sequence-stable separation bound. Section A derives the required uniform separation from primitives under additional limit conditions.

(a) For each n , $\theta \in \Theta$, and shock realization ε , the Picard map $\Gamma_{\theta,n}(y) := h_{\theta,n}(y, X_n) + \varepsilon$ has a unique fixed point $Y_n^\theta(\varepsilon)$ satisfying

$$\|Y_n^\theta(\varepsilon)\|_\infty \leq \frac{M_h + \|\varepsilon\|_\infty}{1 - \rho}.$$

(b) If in addition (C3) holds, then for fixed n and fixed ε ,

$$\|Y_n^\theta(\varepsilon) - Y_n^{\theta'}(\varepsilon)\|_\infty \leq \frac{L_\theta}{1 - \rho} \|\theta - \theta'\|.$$

Proof. See Appendix A.

Part (a) ensures that the equilibrium magnitude is controlled by the shock magnitude, with the contraction factor absorbing the amplification. Part (b) is the parametric sensitivity bound that underlies Lipschitz dependence of the minimum-distance criterion on θ .¹⁸

Local approximation using influence decay. The truncation step controls approximation bias by replacing each ego-neighborhood outcome vector with a local proxy that depends only on nearby shocks. The construction proceeds in two steps. First, Picard iterates propagate information at finite speed through the graph. Second, contraction bounds the gap between any finite iterate and the equilibrium.

Lemma 7 (Local truncation). *Fix a focal node $u \in V_n$, an ego radius $k \geq 1$, and a truncation depth $r \geq 0$. Let ε and ε' be two shock vectors agreeing on $B_{n,k+r}(u)$:*

$$\varepsilon_v = \varepsilon'_v \quad \text{for all } v \in B_{n,k+r}(u).$$

Let $Y_n^\theta(\varepsilon)$ and $Y_n^\theta(\varepsilon')$ be the corresponding equilibria at the same θ . Under (A2), (B1), (B2), and (C2),

$$\|Y_n^\theta(\varepsilon) - Y_n^\theta(\varepsilon')\|_{\infty, B_{n,k}(u)} \leq C_0 \rho^{r+1} (1 + \|\varepsilon\|_\infty + \|\varepsilon'\|_\infty), \quad (3.1)$$

where $C_0 := (2M_h + 1)/(1 - \rho)$.

Proof. See Appendix A.

¹⁸The sensitivity propagates from the structural map through the equilibrium fixed point with amplification factor $(1 - \rho)^{-1}$, which diverges as $\rho \uparrow 1$.

Lemma 7 quantifies how contraction and locality control the effect of shock changes outside $B_{n,k+r}(u)$ on equilibrium outcomes inside $B_{n,k}(u)$.¹⁹ The truncation bound on equilibrium outcomes lifts immediately to a bound on the adversarial loss through the Lipschitz structure of the adaptive test function.

Lemma 8 (Truncation error in the adversarial loss). *Fix a focal node $u \in V_n$ and parameters (θ, ϕ) and recall the per-focal node loss $\ell_{\theta,\phi}(u)$ defined above. Construct a truncated coupling as follows. For each focal node u and depth $r \geq 0$, let $\varepsilon^{(u,\text{out})}$ be an independent copy of the full shock vector ε conditional on (G_n, X_n) , independent across u and independent of $(\varepsilon, \tilde{\varepsilon})$, and define the truncated shock vector*

$$\varepsilon^{(u,r)} := \varepsilon \mathbf{1}_{B_{n,k+r}(u)} + \varepsilon^{(u,\text{out})} \mathbf{1}_{B_{n,k+r}(u)^c}.$$

Define $\tilde{\varepsilon}^{(u,r)}$ analogously using $\tilde{\varepsilon}$ on $B_{n,k+r}(u)$ and an independent copy $\tilde{\varepsilon}^{(u,\text{out})}$ on $B_{n,k+r}(u)^c$. Let $\ell_{\theta,\phi}^{(u,r)}(u)$ be the loss computed from the truncated observed and simulated ego objects generated by $(\varepsilon^{(u,r)}, \tilde{\varepsilon}^{(u,r)})$.

Under (A2), (B1), (B2), (C2), and the adaptive test function regularity in (C6)(i) (including the clipping level η), there exists $C_1 < \infty$ such that for all u, θ, ϕ ,

$$|\ell_{\theta,\phi}(u) - \ell_{\theta,\phi}^{(u,r)}(u)| \leq C_1 \rho^{r+1} (1 + \|\varepsilon\|_\infty + \|\varepsilon^{(u,\text{out})}\|_\infty + \|\tilde{\varepsilon}\|_\infty + \|\tilde{\varepsilon}^{(u,\text{out})}\|_\infty). \quad (3.2)$$

Proof. See Appendix A.

Independence via blocking partition. The blocking step controls dependence-induced variance. The truncated losses $\ell_{\theta,\phi}^{(u,r)}(u)$ are not independent across focal nodes because the truncation balls $B_{n,k+r}(u)$ overlap. To extract independent summands, we construct a dependence graph and exploit the bounded degree structure of the network to partition the focal nodes into a bounded number of independent blocks.

Lemma 9 (Blocking and conditional independence). *Define the dependence graph $H_{n,r} = (V_n, E_H)$ by connecting $u \sim v$ whenever $B_{n,k+r}(u) \cap B_{n,k+r}(v) \neq \emptyset$. Let $\Delta_{n,r}$ denote the maximum degree of $H_{n,r}$. Then:*

(a) $\Delta_{n,r} \leq \nu_n(2(k+r)) - 1$, and the chromatic number of $H_{n,r}$ satisfies $\chi(H_{n,r}) \leq \nu_n(2(k+r))$.

¹⁹The bound decays at rate ρ^{r+1} and grows linearly in the shock maxima, which motivates the truncation sequence r_n in theorem 2.

(b) Under (C4), there exists a partition $V_n = C_{n,r,1} \cup \dots \cup C_{n,r,q_{n,r}}$ with $q_{n,r} \leq \nu_n(2(k+r))$ such that, conditional on (G_n, X_n) , the truncated losses $\{\ell_{\theta,\phi}^{(u,r)}(u) : u \in C_{n,r,j}\}$ are independent for each fixed block j and each fixed (θ, ϕ) .

Proof. See Appendix A.

The blocking construction reduces the dependence created by overlapping ego neighborhoods by partitioning focal nodes into blocks with disjoint truncation balls. The number of blocks satisfies $q_{n,r} \leq \nu_n(2(k+r))$. Concentration bounds for the resulting dependence graph depend on the chromatic number through an effective number of independent local observations

$$m_{\text{eff},n}(r) := \frac{n}{q_{n,r}} \geq \frac{n}{\nu_n(2(k+r))}. \quad (3.3)$$

Consistency requires $m_{\text{eff},n}(r_n) \rightarrow \infty$.²⁰

Remark. When $\nu_n(2(k+r))$ is large, the effective number of independent local observations in (3.3) can be small even when n is large. The role of ball growth in the choice of r_n is revisited in Section 5.1.

Uniform convergence. With bias and dependence controls in hand, we establish a uniform law of large numbers (LLN) for the empirical minimum-distance criterion. The argument has two stages: first, uniform convergence of the truncated criterion via concentration within independent blocks; second, removal of truncation to return to the original criterion.

Lemma 10 (Uniform LLN for the truncated criterion). *Define the truncated empirical and population criteria*

$$\widehat{V}_{n,k}^{(r)}(\theta, \phi) := \frac{1}{n} \sum_{u \in V_n} \ell_{\theta,\phi}^{(u,r)}(u), \quad V_{n,k}^{(r)}(\theta, \phi) := \mathbb{E}[\widehat{V}_{n,k}^{(r)}(\theta, \phi) \mid G_n, X_n].$$

Under (B1), (B2), (C1), (C3'), (C4), and (C6), if $m_{\text{eff},n}(r_n) \rightarrow \infty$ and the entropy condition

$$\frac{2 \log N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) + \log N(\epsilon/L_n, \Theta, \|\cdot\|)}{m_{\text{eff},n}(r_n)} \rightarrow 0 \quad (3.4)$$

²⁰By the lower bound in (3.3), it is sufficient that $\nu_n(2(k+r_n)) = o(n)$.

holds for each fixed $\epsilon > 0$, then

$$\sup_{\theta \in \Theta, \phi \in \Phi_n} |\widehat{V}_{n,k}^{(r_n)}(\theta, \phi) - V_{n,k}^{(r_n)}(\theta, \phi)| \xrightarrow{p} 0.$$

Proof. See Appendix A.

Lemma 11 (Removing the truncation). *Under (A2), (B1), (B2), (C2), (C4), and the adaptive test function regularity in (C6)(i) (including the clipping level η),*

$$\sup_{\theta \in \Theta, \phi \in \Phi_n} |\widehat{V}_{n,k}(\theta, \phi) - \widehat{V}_{n,k}^{(r_n)}(\theta, \phi)| \xrightarrow{p} 0, \quad \sup_{\theta \in \Theta, \phi \in \Phi_n} |V_{n,k}(\theta, \phi) - V_{n,k}^{(r_n)}(\theta, \phi)| \rightarrow 0.$$

Proof. See Appendix A.

Combining lemmas 10 and 11 by the triangle inequality, and passing from the joint criterion to the concentrated minimum-distance criterion using the elementary inequality $|\sup_{\phi} f(\phi) - \sup_{\phi} g(\phi)| \leq \sup_{\phi} |f(\phi) - g(\phi)|$, we obtain the uniform convergence result needed for the Wald argument:

Corollary 2 (Uniform convergence of the concentrated minimum-distance criterion). *Under (A2), (B1), (B2), (C1), (C2), (C3), (C4), and (C6), if $m_{\text{eff},n}(r_n) \rightarrow \infty$ and the entropy condition (3.4) holds,*

$$\sup_{\theta \in \Theta} |\widehat{V}_{n,k}^{\Phi_n}(\theta) - V_{n,k}^{\Phi_n}(\theta)| \xrightarrow{p} 0.$$

Proof. See Appendix A.

3.4 Main result

We now assemble the preceding components into the consistency theorem. The argument combines compactness of Θ , uniform separation of the population criterion, and uniform convergence of the concentrated empirical minimum-distance criterion to show that approximate minimizers converge to the true parameter.

Theorem 2 (Consistency). *Fix an ego radius $k \geq 1$. Let $\hat{\theta}_{n,k}$ be any estimator satisfying the approximate argmin condition*

$$\widehat{V}_{n,k}^{\Phi_n}(\hat{\theta}_{n,k}) \leq \inf_{\theta \in \Theta} \widehat{V}_{n,k}^{\Phi_n}(\theta) + \xi_n, \quad \xi_n \xrightarrow{p} 0.$$

Suppose assumptions (A1), (A2), (B1), and (B2) from Section 2 and assumptions (C1), (C2), (C3), (C4), (C5), (C6), and (C7) hold and that Proposition 1 holds. Assume further that there exists a truncation sequence $r_n \uparrow \infty$ satisfying:

(i) Truncation decay: $\rho^{r_n} \|\varepsilon_n\|_\infty \xrightarrow{p} 0$ and $\rho^{r_n} \|\tilde{\varepsilon}_n\|_\infty \xrightarrow{p} 0$;

(ii) Effective sample growth: $m_{\text{eff},n}(r_n) \rightarrow \infty$ and the entropy condition (3.4) holds.

Then $\hat{\theta}_{n,k} \xrightarrow{p} \theta_0$.

Proof. See Appendix A.

Intuitively, theorem 2 is an argmin-consistency (Wald-type) result for dependent network data. Corollary 2 gives uniform stochastic equicontinuity of the empirical concentrated criterion around its sieve-population analogue, and (C7) controls sieve approximation bias relative to the full population criterion function $V_{n,k}^*$. Proposition 1 supplies the key identification inequality: for each $\epsilon > 0$, parameters with $\|\theta - \theta_0\| \geq \epsilon$ incur at least $\delta(\epsilon) > 0$ additional population loss for all sufficiently large n . With vanishing optimization error ξ_n , an approximate empirical minimizer cannot stay outside an ϵ -neighborhood of θ_0 , hence $\hat{\theta}_{n,k} \xrightarrow{p} \theta_0$.

The truncation sequence r_n is the key tuning bridge between theory and practice. Larger r_n reduces truncation bias through ρ^{r_n} , but it also enlarges overlap neighborhoods and can reduce $m_{\text{eff},n}(r_n)$. The theorem formalizes that, when this bias-dependence tradeoff is balanced and sieve approximation error vanishes, the estimator converges in probability to θ_0 .

4 Estimation Algorithm

Section 3 establishes consistency for approximate minimizers of the concentrated criterion. This section describes how such approximate minimizers are computed in finite samples. Section 4.1 presents an alternating minimax algorithm for approximating

$$\inf_{\theta \in \Theta} \sup_{\phi \in \Phi_n} \widehat{V}_{n,k}(\theta, \phi).$$

Section 4.2 discusses implementation details that matter for finite-sample econometric performance: stabilization of stochastic updates and score computation through equilibrium iteration. Section 4.3 discusses sieve choice as the finite-sample counterpart of assumptions (C6) and (C7).

Throughout this section, the notation and assumptions of Sections 2 and 3 remain in force. In particular, (G_n, X_n) is conditioned on and treated as fixed, $\{D_\phi : \phi \in \Phi_n\}$ denotes the test-function sieve, $\widehat{V}_{n,k}(\theta, \phi)$ denotes the empirical criterion, and $\widehat{V}_{n,k}^{\Phi_n}(\theta) := \sup_{\phi \in \Phi_n} \widehat{V}_{n,k}(\theta, \phi)$ denotes the concentrated sieve minimum-distance criterion. The specialization to the linear in means model is deferred to Section 5.

4.1 Algorithm

The estimator $\hat{\theta}_{n,k}$ is computed as an approximate minimizer of

$$\widehat{V}_{n,k}^{\Phi_n}(\theta) = \sup_{\phi \in \Phi_n} \widehat{V}_{n,k}(\theta, \phi).$$

For each candidate θ , exact evaluation of the concentrated criterion requires solving the inner supremum over ϕ , which is computationally costly. We therefore use alternating stochastic updates: gradient ascent in ϕ for fixed θ , followed by gradient descent in θ for fixed ϕ . This follows the minimax optimization template of Goodfellow et al. (2014, Section 3, Algorithm 1), adapted to structural estimation by Kaji, Manresa, and Pouliot (2023). We refer to the two phases as the adaptive test function phase and the structural-parameter phase.

The algorithm requires three computational primitives. First, an equilibrium solver: for each candidate $\theta \in \Theta$ and shock realization $\tilde{\varepsilon}_n$, the solver computes the equilibrium \tilde{Y}_n^θ by Picard iteration of the map $y \mapsto h_{\theta,n}(y, X_n) + \tilde{\varepsilon}_n$, exploiting the contraction guaranteed by (B1). Second, an ego extraction routine: given an outcome vector $y \in \mathbb{R}^n$ and a focal node $u \in V_n$, it constructs a labeled representative

$$\tilde{S}_{n,k}(u; y) := (G_n[B_{n,k}(u)], (X_{n,j}, y_j)_{j \in B_{n,k}(u)}, u),$$

and then treats the ego object as its focal-node-preserving label-invariant representation $S_{n,k}(u; y) \in \mathcal{S}_{n,k}$ as defined in section 2.1. Third, a parametric adaptive test function family $\{D_\phi : \phi \in \Phi_n\}$ mapping the ego object space $\mathcal{S}_{n,k}$ to $[\eta, 1 - \eta]$ for a fixed clipping level $\eta \in (0, 1/2)$, as required by (C6).

Both phases operate on minibatches. At each step, a batch of m focal nodes u_1, \dots, u_m is drawn uniformly from V_n . The observed ego objects $S_{n,k}^{\theta_0}(u_i)$ are constructed from the observed outcomes Y_n , and the simulated ego objects $\tilde{S}_{n,k}^\theta(u_i)$ are

constructed from a freshly simulated equilibrium \tilde{Y}_n^θ . The minibatch criterion is

$$\widehat{V}_{n,k}^{(m)}(\theta, \phi) := \frac{1}{m} \sum_{i=1}^m \left\{ \log D_\phi(S_{n,k}^{\theta_0}(u_i)) + \log(1 - D_\phi(\tilde{S}_{n,k}^\theta(u_i))) \right\}. \quad (4.1)$$

This is an unbiased estimator of the full empirical criterion $\widehat{V}_{n,k}(\theta, \phi)$ conditional on the equilibrium draws, provided the focal nodes are sampled uniformly.

The adaptive test function phase performs $n_{\text{disc}} \geq 1$ gradient ascent steps on ϕ per structural-parameter step, corresponding to the choice $k = n_{\text{disc}}$ in Goodfellow et al. (2014, Section 3, Algorithm 1). The structural-parameter phase draws fresh shocks $\tilde{\varepsilon}'_n$, computes an independent equilibrium $\tilde{Y}_n^{\theta'}$, and updates θ by gradient descent. In the original minimax formulation, the structural-parameter loss at fixed ϕ is $\frac{1}{m} \sum_{i=1}^m \log(1 - D_\phi(\tilde{S}_{n,k}^\theta(u_i)))$, which can yield weak gradients when D_ϕ nearly separates observed and simulated ego objects.²¹ Following Goodfellow et al. (2014, Section 3, discussion immediately after Eq. (1)), we replace the structural-parameter loss with the nonsaturating variant

$$L_G(\theta; \phi) := -\frac{1}{m} \sum_{i=1}^m \log D_\phi(\tilde{S}_{n,k}^\theta(u'_i)), \quad (4.2)$$

whose gradient with respect to the outcome coordinates y_B of the ego object remains informative when $D_\phi(s)$ is near zero.²²

This modification changes the structural-parameter update rule but not the population target. For each θ , the adaptive test function best response is

$$D_\theta^*(s) = \frac{dP_{n,k}^{\theta_0}}{d(P_{n,k}^{\theta_0} + P_{n,k}^\theta)}(s),$$

so at $\theta = \theta_0$ we have $D_{\theta_0}^*(s) \equiv 1/2$ and $V_{n,k}^*(\theta_0) = -\log 4$.²³

²¹If $D_\phi(s) = \sigma(\ell_\phi(s))$ with $\sigma(t) := (1 + e^{-t})^{-1}$, then $\nabla_{y_B} \log(1 - D_\phi(s)) = -D_\phi(s) \nabla_{y_B} \ell_\phi(s)$, which vanishes as $D_\phi(s) \downarrow 0$. Here ∇_{y_B} differentiates with respect to the continuous outcome coordinates $y_B := (y_j)_{j \in B_{n,k}(u)}$ of the ego object, holding discrete components fixed.

²²Its local derivative is $\nabla_{y_B} (-\log D_\phi(s)) = -\nabla_{y_B} D_\phi(s) / D_\phi(s)$. Under the logit parameterization $D_\phi(s) = \sigma(\ell_\phi(s))$, this equals $-(1 - D_\phi(s)) \nabla_{y_B} \ell_\phi(s)$, so it remains of order $\|\nabla_{y_B} \ell_\phi(s)\|$ near $D_\phi(s) = 0$. The structural-parameter gradient $\nabla_\theta L_G(\theta; \phi)$ composes this derivative with $\partial_\theta(\tilde{Y}_n^\theta|_{B_{n,k}(u)})$ from Section 4.2.

²³Accordingly, the nonsaturating loss (4.2) affects the optimization path but not the identification and consistency characterizations of the minimizer of $V_{n,k}^*(\theta)$ in Sections 2 and 3.

The algorithm is summarized in Algorithm 1. The output $\hat{\theta}_{n,k} := \theta^{(N_{\text{steps}})}$ serves as an approximate minimizer of $\widehat{V}_{n,k}^{\Phi_n}(\theta)$.²⁴ Theorem 2 guarantees consistency provided the approximate argmin condition $\widehat{V}_{n,k}^{\Phi_n}(\hat{\theta}_{n,k}) \leq \inf_{\theta} \widehat{V}_{n,k}^{\Phi_n}(\theta) + \xi_n$ holds with $\xi_n \xrightarrow{p} 0$.

Algorithm 1 Adversarial structural estimation

Require: Observed data (G_n, X_n, Y_n) ; ego radius $k \geq 1$; batch size m ; adaptive test function steps n_{disc} ; learning rates $\eta_D, \eta_G > 0$; iteration count N_{steps} ; Picard tolerance $\tau > 0$; iteration bound T_{max} .

Ensure: $\hat{\theta}_{n,k}$

- 1: **Precompute:** For each $u \in V_n$, store $B_{n,k}(u)$, $G_n[B_{n,k}(u)]$, and $(X_{n,j})_{j \in B_{n,k}(u)}$.
 - 2: **Initialize:** $\theta^{(0)} \in \Theta$, $\phi^{(0)} \in \Phi_n$.
 - 3: **for** $s = 1, \dots, N_{\text{steps}}$ **do**
 - 4: **Simulate equilibrium:** draw $\tilde{\varepsilon}_n \sim f_{\varepsilon}^{\otimes n}$ independently of ε_n .
 - 5: $y^{(0)} \leftarrow 0$
 - 6: **for** $t = 0, 1, \dots, T_{\text{max}} - 1$ **do**
 - 7: $y^{(t+1)} \leftarrow h_{\theta^{(s-1)}, n}(y^{(t)}, X_n) + \tilde{\varepsilon}_n$
 - 8: **if** $\|y^{(t+1)} - y^{(t)}\|_{\infty} < \tau$ **then**
 - 9: **break**
 - 10: **end if**
 - 11: **end for**
 - 12: $\tilde{Y}_n^{\theta^{(s-1)}} \leftarrow y^{(t+1)}$
 - 13: **Adaptive test function phase:**
 - 14: **for** $t = 1, \dots, n_{\text{disc}}$ **do**
 - 15: Draw $u_1, \dots, u_m \sim \text{Unif}(V_n)$ independently.
 - 16: Construct $(S_{n,k}^{\theta_0}(u_i))_{i=1}^m$ from Y_n and $(\tilde{S}_{n,k}^{\theta^{(s-1)}}(u_i))_{i=1}^m$ from $\tilde{Y}_n^{\theta^{(s-1)}}$.
 - 17: $\phi \leftarrow \phi + \eta_D \nabla_{\phi} \widehat{V}_{n,k}^{(m)}(\theta^{(s-1)}, \phi)$ ▷ (4.1)
 - 18: **end for**
 - 19: **Structural-parameter phase:**
 - 20: Draw $\tilde{\varepsilon}'_n \sim f_{\varepsilon}^{\otimes n}$ and compute $\tilde{Y}_n^{\theta^{(s-1)'}}$ by Picard iteration as above.
 - 21: Draw $u'_1, \dots, u'_m \sim \text{Unif}(V_n)$ independently.
 - 22: Construct $(\tilde{S}_{n,k}^{\theta^{(s-1)'}}(u'_i))_{i=1}^m$ from $\tilde{Y}_n^{\theta^{(s-1)'}}$.
 - 23: $\theta \leftarrow \theta - \eta_G \nabla_{\theta} L_G(\theta; \phi)$ ▷ (4.2)
 - 24: **end for**
 - 25: **return** $\hat{\theta}_{n,k} \leftarrow \theta^{(N_{\text{steps}})}$
-

²⁴Within each outer iteration, the adaptive test function phase reuses the same simulated equilibrium draw $\tilde{Y}_n^{\theta^{(s-1)}}$ across the n_{disc} updates, reducing Monte Carlo variance at the cost of conditional dependence across those updates. In the structural-parameter phase, fresh shocks $\tilde{\varepsilon}'_n$ produce an independent equilibrium $\tilde{Y}_n^{\theta^{(s-1)'}}$, which typically stabilizes alternating updates by reducing adaptation to a specific simulation draw.

4.2 Training mechanics

Implementing Algorithm 1 amounts to stochastic optimization of a nonconvex criterion with dependent local observations. For econometric purposes, two implementation choices are central: (i) how finite-sample instability is regularized during training, and (ii) how the structural score with respect to θ is computed. We discuss these issues in turn.

Finite-sample stabilization. On a finite graph with n vertices, the empirical criterion $\widehat{V}_{n,k}(\theta, \phi)$ is a finite average over n ego objects whose neighborhoods overlap. Three sources of instability arise in this setting, each addressed by a corresponding device.

Memorization through overlap. Each vertex $v \in V_n$ appears in $|B_{n,k}^{-1}(v)| := |\{u \in V_n : v \in B_{n,k}(u)\}|$ distinct ego objects. Under bounded degree, this count is at most $\nu_n(k)$.²⁵

To mitigate memorization, the adaptive test function inputs are perturbed by additive noise. For each ego object, the outcome coordinates are perturbed as $y_j \leftarrow y_j + \sigma_{\text{noise}} \cdot z_j$ with $z_j \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, where $\sigma_{\text{noise}} > 0$ is a noise scale. The same perturbation is applied to both observed and simulated ego objects, so that the adaptive test function operates on smoothed versions of both distributions. As training progresses and the simulated distribution approaches the observed distribution, the noise scale is annealed toward zero according to a deterministic schedule $\sigma_{\text{noise}}(s)$ indexed by the outer iteration s . The annealing schedule is a tuning parameter; the implementation uses $\sigma_{\text{noise}}(s) = \sigma_0 \cdot \max(1 - s/s_{\text{anneal}}, 0)$ for constants $\sigma_0 > 0$ and $s_{\text{anneal}} \in \{1, \dots, N_{\text{steps}}\}$.

This perturbation modifies the effective criterion by convolving both the observed and simulated distributions of local observations with a Gaussian kernel on the outcome coordinates: it replaces $P_{n,k}^{\theta_0}$ and $P_{n,k}^{\theta}$ by $P_{n,k}^{\theta_0} * \mathcal{N}(0, \sigma_{\text{noise}}^2 I)$ and $P_{n,k}^{\theta} * \mathcal{N}(0, \sigma_{\text{noise}}^2 I)$, respectively. For $\sigma_{\text{noise}} > 0$, both smoothed distributions admit Lebesgue densities, which ensures that their supports overlap. Without this overlap, the Jensen–Shannon divergence approaches its maximum of $\log 2$ and the optimal adaptive test function achieves nearly perfect separation, rendering its gradient uninformative for

²⁵An adaptive test function with parameter count exceeding the number of distinct ego objects can memorize vertex-specific patterns instead of learning distributional contrasts between observed and simulated ego objects. As in Goodfellow et al. (2014, Section 3), full optimization on finite data can overfit; network overlap amplifies this because a single vertex appears in many ego objects.

the structural parameter. The additive noise prevents this degeneracy by keeping the Jensen–Shannon divergence below $\log 2$ and maintaining an informative adaptive test function gradient throughout training. As $\sigma_{\text{noise}}(s) \rightarrow 0$, the perturbed criterion converges to the original criterion $\widehat{V}_{n,k}(\theta, \phi)$.

Gradient instability. The structural-parameter gradient passes through the equilibrium map, which amplifies perturbations by a factor of $(1 - \rho)^{-1}$ as shown in lemma 6(b). When ρ is close to one, this amplification can produce large gradient norms that destabilize the optimization. To control this, the structural-parameter gradient is clipped to a maximum Euclidean norm: if $\|g\| > C_{\text{clip}}$ for a gradient vector $g := \nabla_{\theta} L_G(\theta; \phi)$, the update uses $C_{\text{clip}} \cdot g / \|g\|$ in place of g . The clipping threshold $C_{\text{clip}} > 0$ is a tuning parameter. This modification preserves the gradient direction while bounding the step size, and does not affect the fixed points of the optimization.

Batch overlap. When focal nodes u_1, \dots, u_m are drawn independently and uniformly from V_n , the corresponding ego neighborhoods $B_{n,k}(u_1), \dots, B_{n,k}(u_m)$ may overlap, particularly around high degree vertices. Overlap within a minibatch concentrates the batch on a shared set of vertices and reduces the effective diversity of the gradient signal.²⁶

The consistency result of Section 3 (theorem 2) concerns the full empirical criterion $\widehat{V}_{n,k}^{\Phi_n}(\theta) = \sup_{\phi \in \Phi_n} \widehat{V}_{n,k}(\theta, \phi)$, not the minibatch approximation used by the optimizer.²⁷

Differentiation through equilibrium iteration. The structural-parameter phase of Algorithm 1 requires the gradient $\nabla_{\theta} L_G(\theta; \phi)$, which depends on θ through the simulated equilibrium \tilde{Y}_n^{θ} . This section describes how this gradient is computed by automatic differentiation through the unrolled Picard iteration.

Fix $\theta \in \Theta$ and a shock realization $\tilde{\varepsilon}_n$. The Picard iteration defines a sequence of outcome vectors $y^{(0)} := 0$, $y^{(t+1)} := h_{\theta,n}(y^{(t)}, X_n) + \tilde{\varepsilon}_n$ for $t = 0, 1, \dots$. The iteration

²⁶A simple fix is greedy packing: draw focal nodes sequentially and accept candidate u only if $B_{n,k}(u) \cap B_{n,k}(u_i) = \emptyset$ for previously accepted u_i . Under bounded degree, a greedy packing with disjoint radius- k balls contains at least $n/\nu_n(2k)$ nodes, so batches are fillable whenever $m \leq n/\nu_n(2k)$.

²⁷The full criterion averages over all focal nodes and overlaps; lemmas 9 to 11 establish its uniform convergence. Packed minibatches only improve stochastic-gradient quality by reducing within-batch dependence and do not create dependence structures beyond those already handled by the blocking argument.

is terminated at a finite index $T = T(\theta, \tilde{\varepsilon}_n)$ determined by the stopping criterion $\|y^{(T+1)} - y^{(T)}\|_\infty < \tau$ or $T = T_{\max}$. The truncated iterate $y^{(T)}$ serves as a computable approximation to the true equilibrium $\tilde{Y}_n^\theta = \lim_{t \rightarrow \infty} y^{(t)}$.

Each iterate $y^{(t+1)} = h_{\theta,n}(y^{(t)}, X_n) + \tilde{\varepsilon}_n$ is a composition of differentiable maps in θ under assumption (D1). The truncated iterate $y^{(T)}$ is therefore a T -fold composition of differentiable maps, and its derivative $\partial_\theta y^{(T)}$ can be computed exactly by applying the chain rule through the composition. Automatic differentiation frameworks record the computation graph of the Picard iteration and compute $\partial_\theta y^{(T)}$ by reverse mode differentiation, without requiring explicit construction of the Jacobian matrices.

The approximation error between the truncated derivative and the derivative of the true equilibrium is controlled by the contraction rate. By lemma 1(b), $\|y^{(T)} - \tilde{Y}_n^\theta\|_\infty \leq \rho^T \|y^{(0)} - \tilde{Y}_n^\theta\|_\infty$, so the equilibrium approximation converges geometrically. For the derivative, implicit differentiation of the fixed point equation $\tilde{Y}_n^\theta = h_{\theta,n}(\tilde{Y}_n^\theta, X_n) + \tilde{\varepsilon}_n$ yields, by Lemma 1',

$$\partial_\theta \tilde{Y}_n^\theta = (I_n - A_{\theta,n}(\tilde{Y}_n^\theta, X_n))^{-1} \partial_\theta h_{\theta,n}(\tilde{Y}_n^\theta, X_n).$$

In the linear case, where $A_{\theta,n}(y, X_n)$ is constant along the Picard path, unrolling T Picard steps corresponds to truncating the Neumann expansion of $(I_n - A_{\theta,n})^{-1}$. In the nonlinear case, the unrolled derivative uses Jacobian products evaluated along the path and still converges to $\partial_\theta \tilde{Y}_n^\theta$ under (D1).²⁸

The contraction assumption (B1) is the property that makes unrolled differentiation viable. Without $\|A_{\theta,n}\|_\infty < 1$, the Picard iteration may not converge, the Neumann series may not converge, and the truncated derivative may not approximate the true derivative.²⁹

The computational cost of equilibrium simulation and reverse-mode differentiation is linear in n per unrolled step under locality.³⁰

²⁸The tail contribution is controlled by $\|A_{\theta,n}(y, X_n)\|_\infty \leq \rho$, so products of t Jacobians are $O(\rho^t)$. With additional uniform Lipschitz regularity of $y \mapsto A_{\theta,n}(y, X_n)$ and $y \mapsto \partial_\theta h_{\theta,n}(y, X_n)$, the path-evaluation component is also geometric, yielding $\|\partial_\theta y^{(T)} - \partial_\theta \tilde{Y}_n^\theta\|_\infty = O(T\rho^T)$. Without that extra regularity, convergence remains qualitative but a uniform geometric rate need not hold.

²⁹In implementations, one enforces contraction by constraining the parameterization of $h_{\theta,n}$ so that $\|A_{\theta,n}(y, X_n)\|_\infty \leq \rho < 1$ for all y and all $\theta \in \Theta$; the concrete constraint is model-specific (see Section 5 for linear in means).

³⁰The forward and reverse passes are both $O(Tn\bar{d})$, where \bar{d} is average degree. Memory is linear in T when the full computation graph is stored. If $\rho \leq \bar{\rho} < 1$, then $T = O(\log(1/\tau)/|\log \rho|)$ under the stopping criterion $\|y^{(T+1)} - y^{(T)}\|_\infty < \tau$, so time and memory scale logarithmically with tolerance.

4.3 Sieve choice

The consistency result of theorem 2 depends on two conditions that constrain the test-function sieve: the entropy condition (C6)(ii), which limits the complexity of the sieve relative to the effective number of independent local observations, and the approximation condition (C7), which requires the sieve to be rich enough to approximate the population optimal adaptive test function. These two conditions impose opposing requirements on the capacity of $\{D_\phi : \phi \in \Phi_n\}$, and their interplay governs the choice of adaptive test function architecture in practice.

The entropy condition requires $\log N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) = o(m_{\text{eff},n})$ for each fixed $\epsilon > 0$, where \mathcal{F}_n is the function class induced by the test-function sieve and $m_{\text{eff},n}$ is the effective number of independent local observations from (3.3).³¹

The approximation condition requires that $\alpha_n := \sup_\theta |V_{n,k}^*(\theta) - V_{n,k}^{\Phi_n}(\theta)| \rightarrow 0$. By lemma 3, the population optimal adaptive test function at parameter θ is $D^*(s) = dP_{n,k}^{\theta_0}/d(P_{n,k}^{\theta_0} + P_{n,k}^\theta)(s)$, so the sieve must approximate this likelihood ratio uniformly over $\theta \in \Theta$.³²

In practice, the adaptive test function should be chosen to reflect the dependence structure implied by the ego object space $\mathcal{S}_{n,k}$. Because ego objects are focal-node-centered graphs with node covariates and outcomes, message-passing architectures are natural candidates for this input format (Xu et al. 2019).³³

More broadly, both the structural model class and the adaptive test function architecture are researcher specified choices, analogous to the choice of moment conditions in the generalized method of moments or the choice of instruments in instrumental variables estimation. A structural model that is too restrictive relative to the data generating process will be misspecified regardless of the adaptive test function, while an adaptive test function that is too restrictive relative to the structural model will fail the approximation condition (C7) and may not detect distributional differences that are present at the population level. Conversely, an overparameterized adaptive test function on a finite graph may satisfy (C7) but violate the entropy condition (C6)(ii),

³¹For parametric families $\Phi_n \subseteq \mathbb{R}^{d_\phi}$ with Lipschitz parameterization and bounded diameter, standard bounds give $\log N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) \leq C_1 d_\phi \log(C_2/\epsilon)$; see Vaart and Wellner (1996) and Vershynin (2025). For fixed ϵ , this reduces (up to constants) to $d_\phi = o(m_{\text{eff},n})$.

³²The complexity of D^* on $\mathcal{S}_{n,k}$ is tied to structural-model complexity: richer local distributional structure implies a richer likelihood-ratio class and requires a more expressive test-function sieve.

³³A useful rule is to use at least k message-passing layers so that the focal-node embedding depends on the full radius- k ball. With fewer layers, the adaptive test function only accesses a strict subset of local information and may fail the approximation condition.

leading to overfitting of the empirical criterion. The researcher must select both components so that the adaptive test function can represent the distributional contrasts implied by the structural model while satisfying the entropy condition (C6)(ii) at the effective number of independent local observations available on the observed graph.

5 Application: Linear-in-Means

This section verifies the performance of our estimator and the associated computational algorithm in a synthetic graph based on the conventional linear-in-means model. This section is split into two parts: Section 5.1 describes the model and synthetic sample, while Section 5.2 assesses the performance of our estimator on it.

5.1 Simulation setup

The structural model takes the form

$$Y_n = \beta W_n Y_n + \gamma X_n + \varepsilon_n, \quad \varepsilon_{n,i} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \quad (5.1)$$

where $G_n = (V_n, E_n)$ is the observed undirected graph on n vertices, W_n is the row-normalized network matrix, X_n is a vector of node covariates, and $\theta = (\beta, \gamma)$ is the structural parameter. The interaction map is

$$h_{\theta,n}(y, X_n) := \beta W_n y + \gamma X_n.$$

Appendix C verifies that this model satisfies the assumptions of Sections 2 and 3 (including the effective Lipschitz condition (C3') used in consistency). The discussion here therefore starts from those verified primitives and concentrates on econometrically relevant implications for identification, rates, implementation, and diagnostics. We first characterize identification in the Gaussian reduced form implied by (5.1), and then translate the consistency conditions of Section 3 into explicit growth restrictions for this model.

The equilibrium (5.1) is linear in ε_n , so

$$Y_n^\theta = (I_n - \beta W_n)^{-1}(\gamma X_n + \varepsilon_n)$$

is multivariate Gaussian with mean and covariance

$$\mu_n^\theta := (I_n - \beta W_n)^{-1} \gamma X_n, \quad \Sigma_n^\theta := (I_n - \beta W_n)^{-1} (I_n - \beta W_n^\top)^{-1}. \quad (5.2)$$

By corollary 1, the marginal law

$$L_u^\theta = \mathcal{L}((Y_{n,j}^\theta)_{j \in B_{n,k}(u)} \mid G_n, X_n)$$

of the outcome subvector indexed by $B := B_{n,k}(u)$ is Gaussian with mean $\mu_{B,\theta,u} := [\mu_n^\theta]_B$ and covariance $\Sigma_{B,\theta,u} := [\Sigma_n^\theta]_{B,B}$, the restrictions of (5.2) to the index set B .

The identification argument of theorem 1 reduces, in the Gaussian case, to local injectivity of the map

$$\theta \mapsto (\mu_{B_{n,k}(u),\theta,u}, \Sigma_{B_{n,k}(u),\theta,u})_{u \in V_n}.$$

The following argument verifies, for the linear Gaussian model, that assumption (D2) is implied by first order variation of the local mean or covariance along the given direction, and characterizes the structure of that variation in terms of the graph.

Write

$$\dot{\mu}_{B,u} := \partial_t [\mu_{B,\theta_0+t\delta,u}]_{t=0}, \quad \dot{\Sigma}_{B,u} := \partial_t [\Sigma_{B,\theta_0+t\delta,u}]_{t=0}.$$

The density $f_{B|\theta,u} = \phi(\cdot; \mu_{B,\theta,u}, \Sigma_{B,\theta,u})$, where $\phi(\cdot; \mu, \Sigma)$ denotes the multivariate Gaussian density with mean μ and covariance Σ , is C^∞ in θ at each fixed $y_B \in \mathbb{R}^{|B|}$. The pointwise θ derivative

$$g_{u,\delta}(y_B) := \partial_t f_{B|\theta_0+t\delta,u}(y_B) \Big|_{t=0} = f_{B|\theta_0,u}(y_B) \cdot s_{u,\delta}(y_B)$$

exists at every y_B , where the score function is

$$\begin{aligned} s_{u,\delta}(y_B) &:= (y_B - \mu_{B,\theta_0,u})^\top \Sigma_{B,\theta_0,u}^{-1} \dot{\mu}_{B,u} - \frac{1}{2} \text{tr}(\Sigma_{B,\theta_0,u}^{-1} \dot{\Sigma}_{B,u}) \\ &\quad + \frac{1}{2} (y_B - \mu_{B,\theta_0,u})^\top \Sigma_{B,\theta_0,u}^{-1} \dot{\Sigma}_{B,u} \Sigma_{B,\theta_0,u}^{-1} (y_B - \mu_{B,\theta_0,u}). \end{aligned} \quad (5.3)$$

Since the score representation (5.3) shows that $s_{u,\delta}$ is a polynomial of degree at most two in $(y_B - \mu_{B,\theta_0,u})$, the function $g_{u,\delta} = f_{B|\theta_0,u} \cdot s_{u,\delta}$ lies in $L^1(\mathbb{R}^{|B|})$, and $\|g_{u,\delta}\|_{L^1} = \mathbb{E}_{Y_B \sim P_{B,\theta_0,u}}[|s_{u,\delta}(Y_B)|]$. When $\dot{\mu}_{B,u} \neq 0$ or $\dot{\Sigma}_{B,u} \neq 0$, the score $s_{u,\delta}$ is a nonzero polynomial in y_B ; the zero set of any nonzero polynomial in $\mathbb{R}^{|B|}$ has Lebesgue measure

zero, and $P_{B,\theta_0,u}$ is absolutely continuous, so $P_{B,\theta_0,u}(s_{u,\delta} = 0) = 0$ and therefore $\|g_{u,\delta}\|_{L^1} > 0$.

For the remainder bound in (D2), the density $f_{B|\theta,u}$ is C^2 in θ by (C.1) and the C^∞ regularity of the Gaussian density. Taylor's theorem in t gives

$$\|f_{B|\theta_0+t\delta,u} - f_{B|\theta_0,u} - t g_{u,\delta}\|_{L^1} \leq \frac{t^2}{2} \sup_{|s|\leq t_0} \|\partial_t^2 f_{B|\theta_0+s\delta,u}\|_{L^1(\mathbb{R}^{|B|})}.$$

The second θ derivative of the Gaussian density is of the form (polynomial in y_B) times (Gaussian kernel in y_B), and hence has finite L^1 norm.³⁴ Choose $t_0 > 0$ such that

$$\frac{t_0}{2} \sup_{|s|\leq t_0} \|\partial_t^2 f_{B|\theta_0+s\delta,u}\|_{L^1} \leq \frac{c_g}{2}$$

yields the bound required by (D2). Accordingly, for the linear Gaussian model, assumption (D2) reduces to the requirement that for each $\delta \in \mathbb{S}^{p-1}$ there is a fraction $c_0 > 0$ of focal nodes u such that

$$(\dot{\mu}_{B_{n,k}(u),u}, \dot{\Sigma}_{B_{n,k}(u),u}) \neq (0, 0) \quad \text{and} \quad \|g_{u,\delta}\|_{L^1} \geq c_g.$$

This is a condition on the graph sequence and covariates. It rules out parameter directions along which the local outcome distribution is uniformly insensitive across a positive fraction of focal nodes. The reduced form representation $\mu_n^\theta = \sum_{t=0}^\infty \beta^t W_n^t \gamma X_n$ shows that the (i, j) entry of $(I_n - \beta W_n)^{-1}$ equals $\sum_{t=0}^\infty \beta^t (W_n^t)_{ij}$, a weighted sum over walks of length t from i to j in G_n . Restricting to the index set $B_{n,k}(u)$, the local mean $\mu_{B,\theta,u}$ and covariance $\Sigma_{B,\theta,u}$ encode the structure of these weighted walk counts within and around the radius k ball of u . Two focal nodes u and v whose induced subgraphs $G_n[B_{n,k}(u)]$ and $G_n[B_{n,k}(v)]$ have distinct topology, or whose covariate configurations differ under every focal-node-preserving relabeling, generate local mean and covariance parameters that respond differently to perturbations of θ . Assumption (D2) formalizes this requirement: a positive fraction of focal nodes must be positioned so that their local outcome distributions vary nontrivially in every parameter direction.

Assumption (D3), which requires the map $u \mapsto \omega_u := \Omega(S_{n,k}^\theta(u))$ to be injective, ensures that focal node-level distributional variation established by (D2) is detectable

³⁴Because $\theta \mapsto \|\partial_t^2 f_{B|\theta,u}\|_{L^1}$ is continuous and Θ is compact, $\sup_{|s|\leq t_0} \|\partial_t^2 f_{B|\theta_0+s\delta,u}\|_{L^1}$ is finite, so the remainder is $O(t^2)$.

in the mixture distribution of local observations. Under (D3), lemma 5 shows that restricting $P_{n,k}^\theta$ to the measurable fiber $\Omega^{-1}(\{\omega_u\})$ identifies the unit-specific law at u , so variation in L_u^θ implies variation of the mixture law. Together, (D2) and (D3) imply local injectivity of $\theta \mapsto P_{n,k}^\theta$ at θ_0 , which by lemmas 3 and 4 makes θ_0 the unique local minimizer of $V_{n,k}^*(\theta)$.

To place these conditions in context, consider a vertex transitive d regular graph with constant covariates $X_{n,i} = x$ for all i . On such a graph, every focal-node-centered radius k ball is isomorphic to every other under a focal-node-preserving relabeling, so all observable signatures ω_u coincide. Assumption (D3) fails: the map $u \mapsto \omega_u$ is not injective, and the mixture deconvolution step of theorem 1 does not apply. In such cases, identification from the k distribution of local observations, if it holds, requires an argument that does not separate unit-specific mixture components via observable signatures. The binding failure condition is (D3). Assumption (D2) is not necessarily violated: for d regular undirected graphs, W_n is symmetric, so the covariance

$$\Sigma_n^\theta = [(I_n - \beta W_n)^{-1}]^2$$

varies nontrivially with β in directions that leave the mean constant, and the score $s_{u,\delta}$ may be nonzero in such directions even when $\dot{\mu}_{B,u} = 0$. Identification fails because (D3) prevents extracting unit-specific distributions of local observations from the mixture, not because (D2) fails.

The identifying variation exploited by the minimum-distance criterion is the positional heterogeneity of focal nodes in the network: different focal nodes are exposed to peer influence through distinct local path structures, and this heterogeneity separates the peer effect parameter β from the exogenous covariate effect γ . The minimum-distance criterion captures this variation through the adaptive test function's access to both local graph topology and local covariate-outcome vectors (X, Y) within each ego object, rather than through a prespecified set of moment restrictions.

Truncation and effective sample size. For $Z \sim \mathcal{N}(0, 1)$, the moment generating function satisfies

$$\mathbb{E}[e^{\lambda|Z|}] = 2e^{\lambda^2/2}\Phi(\lambda) \leq 2e^{\lambda^2/2} < \infty \quad \text{for every } \lambda \in \mathbb{R},$$

where Φ denotes the standard normal distribution function. For Gaussian shocks, a tail bound and union bound over coordinates yield $\|\varepsilon_n\|_\infty = O_p(\sqrt{\log n})$; see the calculation leading to \mathcal{A}_n in Section C.

The truncation decay condition (C4) requires $|\beta|^{r_n}\|\varepsilon_n\|_\infty \rightarrow 0$ in probability, and in the Gaussian case this is satisfied by any sequence $r_n = c \log n$ with $c > 0$.³⁵

Under bounded degree Δ , the ball volume satisfies $\nu_n(2(k+r_n)) \leq \Delta^{2(k+r_n)}$. For $r_n = c \log n$, this gives $\nu_n(2(k+r_n)) \leq \Delta^{2k} \cdot n^{2c \log \Delta}$, and the effective number of independent local observations bound (3.3) yields

$$m_{\text{eff},n}(r_n) \geq \Delta^{-2k} \cdot n^{1-2c \log \Delta}. \quad (5.4)$$

The bound (5.4) implies that $m_{\text{eff},n}(r_n) \rightarrow \infty$ holds whenever $1 - 2c \log \Delta > 0$, that is, whenever $c < 1/(2 \log \Delta)$. For Gaussian shocks, any $c \in (0, 1/(2 \log \Delta))$ simultaneously satisfies the truncation decay requirement and the effective sample growth condition.³⁶

5.2 Implementation and diagnostics

This subsection describes how the three computational primitives and the alternating update scheme of Algorithm 1 take explicit form in the linear model, then links convergence diagnostics to the population criterion benchmark.

Equilibrium solver. The Picard iteration for the linear model is

$$y^{(t+1)} = \beta W_n y^{(t)} + \gamma X_n + \tilde{\varepsilon}_n, \quad y^{(0)} = 0, \quad t = 0, 1, \dots \quad (5.5)$$

Since $A_{\theta,n} = \beta W_n$ does not depend on y , each iterate is an affine map of the previous, and the convergence rate is exactly $|\beta|^t$:

$$\|y^{(T)} - \tilde{Y}_n^\theta\|_\infty \leq |\beta|^T \|\tilde{Y}_n^\theta\|_\infty.$$

The stopping criterion $\|y^{(T+1)} - y^{(T)}\|_\infty < \tau$ is satisfied within $T = O(\log(1/\tau)/|\log|\beta||)$ steps.

³⁵For $|\beta| \in (0, 1)$ and $c > 0$, $|\beta|^{c \log n} \sqrt{\log n} = n^{c \log|\beta|} \sqrt{\log n} \rightarrow 0$ because $c \log|\beta| < 0$.

³⁶An alternative in the Gaussian case is $r_n = c \log \log n$ with $c > 1/|\log|\beta||$, which gives $m_{\text{eff},n}(r_n) \geq n/(\log n)^{2c \log \Delta} \rightarrow \infty$ under bounded degree.

The derivative $\partial_\theta \tilde{Y}_n^\theta$ is computed by backpropagation through the unrolled iteration (5.5). By Lemma 1', $\partial_\theta \tilde{Y}_n^\theta = (I_n - \beta W_n)^{-1} \partial_\theta h_{\theta,n}(\tilde{Y}_n^\theta, X_n)$. Since $A_{\theta,n} = \beta W_n$ is constant along the Picard path, the unrolled derivative of the T step iterate is

$$\partial_\theta y^{(T)} = \sum_{t=0}^{T-1} (\beta W_n)^t \cdot \partial_\theta h_{\theta,n}(y^{(T-1-t)}, X_n), \quad (5.6)$$

a finite truncation of the Neumann series for $(I_n - \beta W_n)^{-1} \partial_\theta h_{\theta,n}$. The approximation error satisfies

$$\|\partial_\theta y^{(T)} - \partial_\theta \tilde{Y}_n^\theta\|_\infty \leq |\beta|^T \|(I_n - \beta W_n)^{-1}\|_\infty \|\partial_\theta h_{\theta,n}(\tilde{Y}_n^\theta, X_n)\|_\infty,$$

decaying at rate $|\beta|^T$.

Ego extraction. For each focal node u and outcome vector $y \in \mathbb{R}^n$, the ego extraction routine returns the triple

$$\tilde{S}_{n,k}(u; y) = (G_n[B_{n,k}(u)], (X_{n,j}, y_j)_{j \in B_{n,k}(u)}, u),$$

which is then interpreted as the focal-node-preserving label-invariant representation $S_{n,k}(u; y) \in \mathcal{S}_{n,k}$.³⁷

Adaptive test function architecture. By lemma 3, the population optimal adaptive test function at parameter θ is

$$D^*(s) = \frac{dP_{n,k}^{\theta_0}}{d(P_{n,k}^{\theta_0} + P_{n,k}^\theta)}(s).$$

For the Gaussian marginal $L_u^\theta = \mathcal{N}(\mu_{B,\theta,u}, \Sigma_{B,\theta,u})$, this ratio is a smooth function of local outcomes and local graph structure through $(I_n - \beta W_n)^{-1}|_B$. This supports message-passing adaptive test functions on ego objects (Xu et al. 2019).³⁸

³⁷Under bounded degree, $|B_{n,k}(u)| \leq \Delta^k$, so extraction cost per focal node is $O(\Delta^k)$ independent of n .

³⁸Using at least k aggregation layers aligns the representation with radius- k ego information and helps satisfy (C7). Whether $\sup_\theta |V_{n,k}^*(\theta) - V_{n,k}^{\Phi_n}(\theta)| \rightarrow 0$ also depends on parameter count and uniform approximation quality over $\theta \in \Theta$.

Alternating updates. The adaptive test function parameters ϕ are updated by $n_{\text{disc}} \geq 1$ gradient ascent steps on the minibatch criterion (4.1) for fixed θ . The structural parameter θ is updated by one gradient descent step on the nonsaturating loss (4.2). For the linear model, the structural-parameter gradient is

$$\nabla_{\theta} L_G(\theta; \phi) = -\frac{1}{m} \sum_{i=1}^m \frac{\nabla_{\theta} D_{\phi}(\tilde{S}_{n,k}^{\theta}(u'_i))}{D_{\phi}(\tilde{S}_{n,k}^{\theta}(u'_i))}, \quad (5.7)$$

where, in (5.7), $\nabla_{\theta} D_{\phi}(\tilde{S}_{n,k}^{\theta}(u))$ is the composition of the test-function Jacobian with the unrolled derivative $\partial_{\theta} \tilde{Y}_n^{\theta}|_{B_{n,k}(u)}$ computed via (5.6).

Finite sample stabilization. The three stabilization devices of Section 4.2 apply directly in the linear-Gaussian setting.³⁹

Convergence diagnostic. Define the discriminator loss as $\mathcal{L}_D(\theta, \phi) := -\widehat{V}_{n,k}^{(m)}(\theta, \phi)$, the negated minibatch criterion from (4.1) that the adaptive test function maximizes, and define the structural-parameter loss as $L_G(\theta; \phi)$ from (4.2), which the structural parameter minimizes. By lemma 3(b) applied with $P = Q = P_{n,k}^{\theta_0}$, the concentrated population minimum-distance criterion at the true parameter satisfies

$$V_{n,k}^*(\theta_0) = -\log 4 + 2 \text{JS}(P_{n,k}^{\theta_0} \parallel P_{n,k}^{\theta_0}) = -\log 4. \quad (5.8)$$

This equality holds for any structural model satisfying (A1), (A2), (B1), and (B2), independently of the specific value of θ_0 and of (G_n, X_n) . At the population optimum, the two distributions of local observations coincide and the optimal adaptive test function is $D^*(s) = 1/2$ for all $s \in \mathcal{S}_{n,k}$: the adaptive test function cannot do better than random classification. Both component expectations in

$$V_{n,k}^*(\theta_0) = \mathbb{E}_{\theta_0}[\log D^*(S_{n,k}^{\theta_0}(U_n))] + \mathbb{E}_{\theta_0}[\log(1 - D^*(S_{n,k}^{\theta_0}(U_n)))]$$

³⁹Additive Gaussian input noise with schedule $\sigma_{\text{noise}}(s) = \sigma_0 \max(1 - s/s_{\text{anneal}}, 0)$ smooths both observed and simulated distributions of local observations while preserving Gaussian form (covariance inflated by $\sigma_{\text{noise}}^2 I_{|B|}$). Gradient clipping at threshold C_{clip} controls amplification by $(1 - |\beta|)^{-1}$ as $|\beta| \uparrow 1$. Disjoint-ball focal-node sampling reduces within-batch overlap, and under bounded degree allows minibatch size $m \leq n/\nu_n(2k) \leq n/\Delta^{2k}$.

equal $\log(1/2)$, giving $\log(1/2) + \log(1/2) = -\log 4$. Consequently, at $D^* \equiv 1/2$, the discriminator loss satisfies

$$\mathcal{L}_D(\theta_0, D^*) = -[\log(1/2) + \log(1/2)] = 2 \log 2,$$

and the structural-parameter loss satisfies $L_G(\theta_0; D^*) = -\log(1/2) = \log 2$.

In an empirical application where θ_0 is unknown, equation (5.8) provides a convergence diagnostic requiring no knowledge of the true parameter.⁴⁰

The tables and figures below summarize one representative simulation.⁴¹ Table 1 reports run level summary statistics for the graph and observed data, and Table 2 reports the final parameter estimates at convergence along with absolute errors relative to the ground truth.

Table 1: Run summary (graph and observed data).

Statistic	Value
Nodes	249,813
Edges (undirected)	688,630
Mean degree	5.513
Max degree	102
Mean Y_{obs}	-0.0078
Std Y_{obs}	1.903
Ordinary least squares (OLS) slope (Y_{obs} on X)	1.545
Picard iterations	15

Notes: The table reports descriptive statistics from the simulated dataset.

Table 2: Final estimates versus truth.

Parameter	True	Estimated	Absolute Error
β	0.4000	0.4096	0.0096
γ	1.5000	1.5046	0.0046

Notes: “Estimated” reports the final iterate of the adversarial estimator. “Absolute Error” is computed as $|\hat{\theta} - \theta_0|$ for each parameter.

⁴⁰At convergence of $\hat{\theta}_{n,k}^{(s)}$ to a local minimizer of $\hat{V}_{n,k}^{\Phi_n}$, one expects both $\hat{V}_{n,k}^{\Phi_n}(\hat{\theta}_{n,k}^{(s)})$ near $-\log 4$ and discriminator outputs on observed/simulated ego objects concentrated near $1/2$. These are equivalent signatures of near-equality of local-observation laws and can be used as stopping criteria.

⁴¹The linear in means equilibrium is simulated on a synthetic graph generated by the Lancichinetti–Fortunato–Radicchi benchmark (Lancichinetti, Fortunato, and Radicchi 2008), and θ is estimated by adversarial training with ego radius $k = 2$.

Figure 1 reports the trajectories of the parameter estimates $\hat{\theta}^{(s)} = (\hat{\beta}^{(s)}, \hat{\gamma}^{(s)})$ as a function of structural-parameter step s , initialized at a point $\theta^{(0)} \neq \theta_0$ and converging to a neighborhood of the true parameter vector θ_0 .

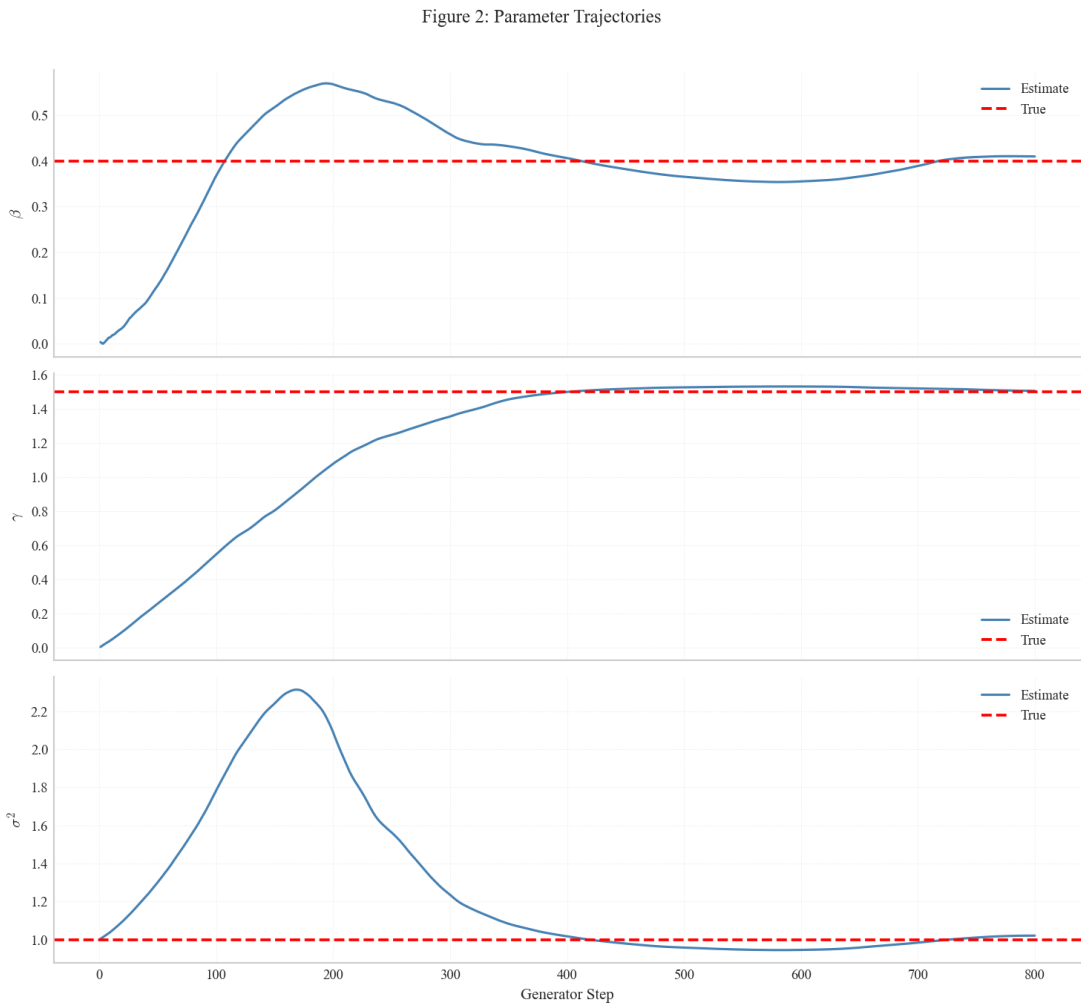


Figure 1: Parameter trajectories $(\hat{\beta}^{(s)}, \hat{\gamma}^{(s)})$ as a function of structural-parameter step s . Horizontal dashed lines indicate the true values β_0 and γ_0 .

Figure 2 plots the discriminator loss $\mathcal{L}_D(\hat{\theta}^{(s)}, \phi^{(s)})$ and structural-parameter loss $L_G(\hat{\theta}^{(s)}; \phi^{(s)})$ over training (rolling mean). At convergence, the discriminator loss stabilizes near $2 \log 2$ and the structural-parameter loss stabilizes near $\log 2$, consistent with discriminator outputs near $1/2$ and hence with $\hat{V}_{n,k}^{\Phi_n}(\hat{\theta}_{n,k}^{(s)})$ near the population optimum $-\log 4$.

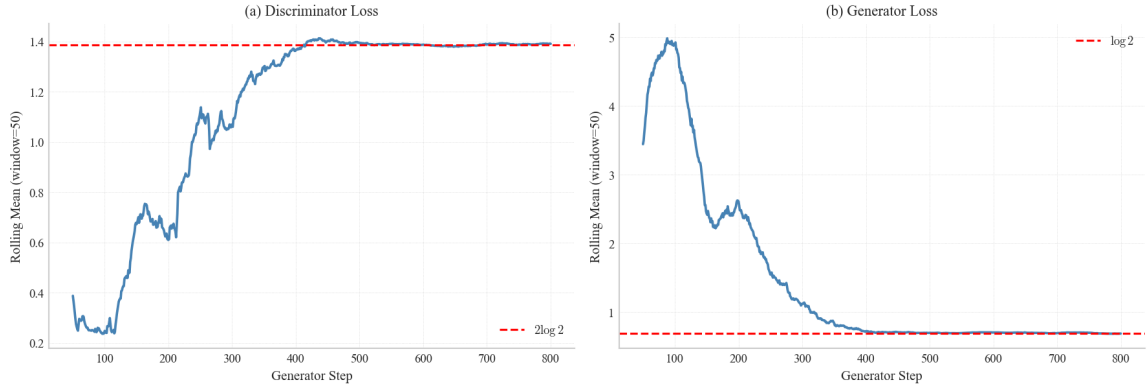


Figure 2: Discriminator loss \mathcal{L}_D and structural-parameter loss L_G trajectories over training (rolling mean). Horizontal dashed lines indicate the asymptotic values $2\log 2$ and $\log 2$ implied by $D^*(\cdot) = 1/2$ at the population optimum.

Figure 3 reports the discriminator score distributions evaluated on held-out observed and simulated ego objects at the final iterate. At convergence, the two distributions overlap near $1/2$, indicating that the adaptive test function cannot reliably distinguish the two distributions of local observations.

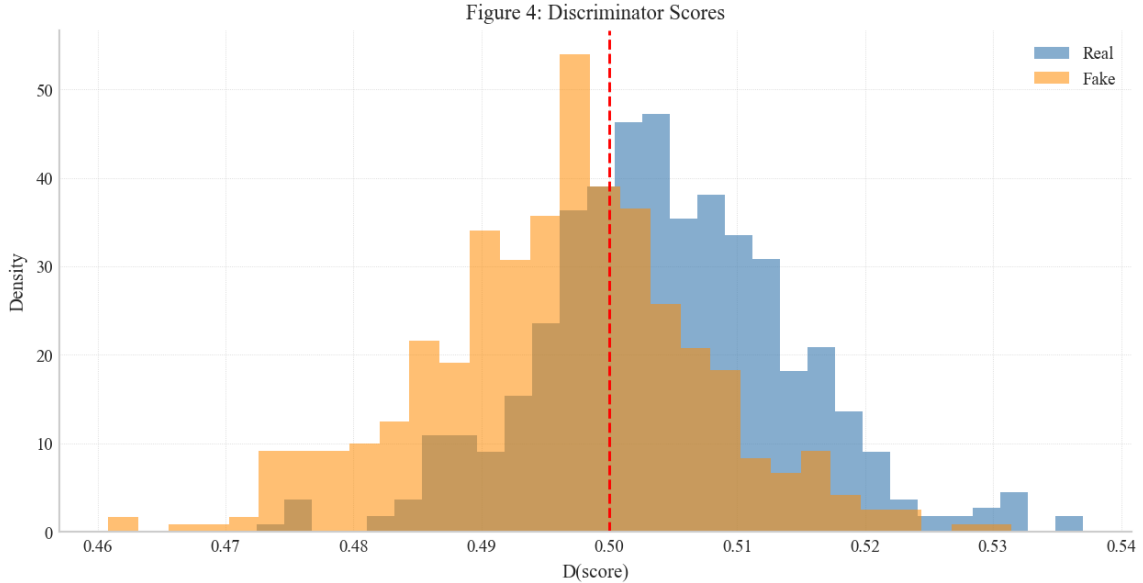


Figure 3: Overlapping histograms of discriminator scores on observed (real) and simulated (fake) ego objects at the final iterate. The vertical dashed line marks $1/2$.

Table 3: Tail statistics over the last 500 structural-parameter steps.

Parameter	Mean (last 500)	Std (last 500)	True
β	0.3886	0.0277	0.4000
γ	1.5048	0.0371	1.5000

Notes: Tail moments are computed from the last 500 structural-parameter iterates of the training path. “True” reports the data-generating parameter values used in the simulation experiment.

To assess tail stability at convergence, Table 3 summarizes the empirical distribution of the parameter iterates over the last 500 structural-parameter steps. The tail means are close to the true values and the within-tail variability is limited, which is consistent with the optimizer remaining in a stable neighborhood of θ_0 at termination.

6 Conclusions

This paper develops an adversarial minimum-distance approach to structural estimation in network equilibrium models when data consist of one observed graph and one observed equilibrium outcome. The core contribution is to combine focal-node-centered local observations with a minimax criterion that adaptively searches over rich test functions, yielding a tractable and interpretable distance between observed and simulated local distributions. On the theory side, the paper establishes population identification and consistency under growing-graph asymptotics with overlap-driven dependence, using local approximation, truncation, and blocking arguments to replace unavailable iid tools. On the computational side, it provides a practical estimation pipeline that aligns discriminator training, equilibrium simulation, and structural updates with the objects appearing in the asymptotic analysis. The linear-in-means specialization illustrates implementation details and confirms stable recovery of structural parameters in a benchmark environment. Taken together, these results show that adversarial structural estimation can be made both theoretically disciplined and computationally feasible in single-graph settings. Future work will focus on formal inference theory beyond bootstrap practice, including adaptations of network-dependence limit-theory tools (Kojevnikov, Marmer, and Song 2021), and on extensions to environments with endogenous network formation, richer strategic heterogeneity, and dynamic or multiple-equilibrium structures.

References

- Arduini, Tiziano, Eleonora Patacchini, and Edoardo Rainone (2015). “Parametric and Semiparametric IV Estimation of Network Models with Selectivity”. EIEF Working Paper 15/09.
- Axler, Sheldon (2020). *Measure, Integration & Real Analysis*. PDF version dated October 27, 2025. Springer.
- Bennett, Andrew, Nathan Kallus, and Tobias Schnabel (2019). “Deep Generalized Method of Moments for Instrumental Variable Analysis”. In: *Advances in Neural Information Processing Systems*. Vol. 32, pp. 3564–3574.
- Blume, Lawrence E. et al. (2015). “Linear Social Interactions Models”. In: *Journal of Political Economy* 123.2, pp. 444–496.
- Bogachev, Vladimir I. (2007). *Measure Theory*. Vol. 1. Springer.
- Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). “Identification of peer effects through social networks”. In: *Journal of Econometrics* 150.1, pp. 41–55.
- (2020). “Peer Effects in Networks: A Survey”. In: *Annual Review of Economics* 12, pp. 603–629.
- Conrad, Keith (Mar. 1, 2025). *Differentiating under the integral sign*.
- Cover, Thomas M. and Joy A. Thomas (2006). *Elements of Information Theory*. 2nd ed. Wiley-Interscience.
- de Paula, Áureo, Imran Rasul, and Pedro Sousa (2023). “Identifying network ties from panel data: theory and an application to tax competition”. In: *Review of Economic Studies*. Forthcoming.
- Drukker, David M., Peter Egger, and Ingmar R. Prucha (2023). “Simultaneous equations models with higher-order spatial or social network interactions”. In: *Econometric Theory* 39.6, pp. 1154–1201.
- Farrell, Max H., Tengyuan Liang, and Sanjog Misra (2021). “Deep Neural Networks for Estimation and Inference”. In: *Econometrica* 89.1, pp. 181–213.
- (2025). “Deep Learning for Individual Heterogeneity: An Automatic Inference Framework”. arXiv:2010.14694v3, revised April 2025.
- Gilmer, Justin et al. (2017). “Neural Message Passing for Quantum Chemistry”. In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. Proceedings of Machine Learning Research. PMLR, pp. 1263–1272.

- Goodfellow, Ian J. et al. (2014). “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems*. Vol. 27. Curran Associates, Inc., pp. 2672–2680.
- Graham, Bryan S. (2015). “Methods of Identification in Social Networks”. In: *Annual Review of Economics* 7, pp. 465–85.
- Hamilton, William L., Rex Ying, and Jure Leskovec (2017). “Inductive Representation Learning on Large Graphs”. In: *Advances in Neural Information Processing Systems*. Vol. 30, pp. 1024–1034.
- Hartford, Jason et al. (2017). “Deep IV: A Flexible Approach for Counterfactual Prediction”. In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. Proceedings of Machine Learning Research. PMLR, pp. 1414–1423.
- Hutchcroft, John (Jan. 24, 2023). *Lecture 5: Complete Metric Spaces and Contraction Mapping Principle*. MIT OpenCourseWare.
- Johnsson, Ida and Hyungsik Roger M. Moon (2021). “Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach”. In: *The Review of Economics and Statistics* 103.2, pp. 328–345.
- Kaji, Tetsuya, Elena Manresa, and Guillaume Pouliot (2023). “An Adversarial Approach to Structural Estimation”. In: *Econometrica* 91.6, pp. 2041–2063.
- Kipf, Thomas N. and Max Welling (2017). “Semi-Supervised Classification with Graph Convolutional Networks”. In: *International Conference on Learning Representations*. Poster.
- Kojevnikov, Denis, Vadim Marmer, and Kyungchul Song (2021). “Limit theorems for network dependent random variables”. In: *Journal of Econometrics* 222.2, pp. 882–908.
- Kranton, Rachel, Martin D’Amours, and Yann Bramoullé (2014). “Strategic Interaction and Networks”. In: *American Economic Review* 104.3, pp. 898–930.
- Lancichinetti, Andrea, Santo Fortunato, and Filippo Radicchi (2008). “Benchmark graphs for testing community detection algorithms”. In: *Physical Review E* 78.4, p. 046110.
- Lebl, Jiri (May 16, 2025). *Basic Analysis: Introduction to Real Analysis, Volume I*.
- Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010). “Specification and estimation of social interaction models with network structures”. In: *The Econometrics Journal* 13.2, pp. 145–176.

- Leung, Michael P. and Pantelis Loupos (2025). “Graph Neural Networks for Causal Inference Under Network Confounding”. arXiv:2211.07823v5, revised December 28, 2025.
- Lewis, Greg and Vasilis Syrgkanis (2018). “Adversarial Generalized Method of Moments”. arXiv:1803.07164, revised April 2018.
- Lin, Xu and Lung-fei Lee (2010). “GMM estimation of spatial autoregressive models with unknown heteroskedasticity”. In: *Journal of Econometrics* 157.1, pp. 34–52.
- Liu, Xiaodong (2014). “Identification and Efficient Estimation of Simultaneous Equations Network Models”. In: *Journal of Business Economics and Statistics* 32.4, pp. 516–536.
- (2020). “GMM identification and estimation of peer effects in a system of simultaneous equations”. In: *Journal of Spatial Econometrics* 1.1. Article 1.
- Liu, Xiaodong and Lung-fei Lee (2010). “GMM estimation of social interaction models with centrality”. In: *Journal of Econometrics* 157.1, pp. 34–52.
- Liu, Xiaodong and Paulo Saraiva (2019). “GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity”. In: *Econometric Reviews* 38.4, pp. 359–385.
- MIT OpenCourseWare (2010). *Example: Change of Variables*. MIT OpenCourseWare.
- Newey, Whitney K. and Daniel McFadden (1994). “Large Sample Estimation and Hypothesis Testing”. In: *Handbook of Econometrics*. Ed. by Robert F. Engle and Daniel L. McFadden. Vol. 4. Elsevier, pp. 2113–2245.
- Ruzhansky, Michael and Mitsuru Sugimoto (2015). “On global inversion of homogeneous maps”. In: *Bulletin of Mathematical Sciences* 5, pp. 13–18.
- Scarselli, Franco et al. (2009). “The Graph Neural Network Model”. In: *IEEE Transactions on Neural Networks* 20.1, pp. 61–80.
- Vaart, Aad W. van der and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.
- Velickovic, Petar et al. (2018). “Graph Attention Networks”. In: *International Conference on Learning Representations*.
- Vershynin, Roman (2025). *High-Dimensional Probability: An Introduction with Applications to Data Science*. 2nd ed. Pre-publication draft dated November 28, 2025. Cambridge University Press.
- Xu, Keyulu et al. (2019). “How Powerful are Graph Neural Networks?” In: *International Conference on Learning Representations*.

A Mathematical proofs

This appendix collects deferred proofs for the identification and consistency results developed in the main text. The ordering follows the sequence in which results appear in Sections 2 and 3.

Proof of Lemma 1. Fix θ , X_n , and ε_n , and define $\Gamma : \mathbb{R}^n \rightarrow \mathbb{R}^n$ by $\Gamma(y) := h_{\theta,n}(y, X_n) + \varepsilon_n$. By (B1), the shock cancels in differences and $\|\Gamma(y) - \Gamma(y')\|_\infty \leq \rho \|y - y'\|_\infty$ with $\rho < 1$, so Γ is a strict contraction on the complete metric space $(\mathbb{R}^n, \|\cdot\|_\infty)$. The Banach fixed point theorem yields a unique fixed point Y_n^θ (Hutchcroft 2023, Theorem 8), establishing (a). The geometric convergence bound in (b) follows from the contraction bound applied to Γ^t .

For (c), initialize the Picard iteration at $Y^{(0)} = 0$. Since $h_{\theta,n}$ is measurable by (B2), $Y^{(1)} = h_{\theta,n}(0, X_n) + \varepsilon_n$ is measurable as a function of ε_n , and by induction so is $Y^{(t)}$ for every t . The limit $Y_n^\theta = \lim_{t \rightarrow \infty} Y^{(t)}$ is measurable because each coordinate is the pointwise limit of measurable real valued functions: for a measurable sequence (f_t) , $\limsup_t f_t = \inf_{N \geq 1} \sup_{t \geq N} f_t$ is measurable as a countable infimum of countable suprema of measurable functions. Since $Y^{(t)} \rightarrow Y_n^\theta$ pointwise by part (b), each coordinate of Y_n^θ equals the corresponding limsup and is therefore measurable.

Proof of Lemma 1'. Fix n and condition on (G_n, X_n) . Under (B2) and (D1), $F_{\theta,n}$ is C^1 with Jacobian $\partial_y F_{\theta,n}(y) = I_n - A_{\theta,n}(y, X_n)$. By (D1)(ii), $\|A_{\theta,n}(y, X_n)\|_\infty \leq \rho < 1$ for all $y \in \mathbb{R}^n$, so the Neumann series converges:

$$(I_n - A_{\theta,n}(y, X_n))^{-1} = \sum_{t=0}^{\infty} A_{\theta,n}(y, X_n)^t, \quad \|(I_n - A_{\theta,n}(y, X_n))^{-1}\|_\infty \leq \frac{1}{1 - \rho}. \quad (\text{A.1})$$

Indeed, letting $S_T := \sum_{t=0}^T A_{\theta,n}(y, X_n)^t$, we have $(I_n - A_{\theta,n}(y, X_n))S_T = I_n - A_{\theta,n}(y, X_n)^{T+1}$. Since $\|A_{\theta,n}(y, X_n)\|_\infty \leq \rho < 1$, $\|A_{\theta,n}(y, X_n)^{T+1}\|_\infty \leq \rho^{T+1} \rightarrow 0$, so S_T converges in operator norm to $(I_n - A_{\theta,n}(y, X_n))^{-1}$ and the bound in (A.1) follows. In particular, $\det \partial_y F_{\theta,n}(y) \neq 0$ for all $y \in \mathbb{R}^n$. Moreover, since $h_{\theta,n}(\cdot, X_n)$ is ρ -Lipschitz in $\|\cdot\|_\infty$ by (B1) and (B2), for every $y \in \mathbb{R}^n$ we have

$$\begin{aligned} \|h_{\theta,n}(y, X_n)\|_\infty &\leq \|h_{\theta,n}(y, X_n) - h_{\theta,n}(0, X_n)\|_\infty + \|h_{\theta,n}(0, X_n)\|_\infty \\ &\leq \rho \|y\|_\infty + \|h_{\theta,n}(0, X_n)\|_\infty. \end{aligned}$$

and therefore

$$\|F_{\theta,n}(y)\|_\infty = \|y - h_{\theta,n}(y, X_n)\|_\infty \geq (1 - \rho)\|y\|_\infty - \|h_{\theta,n}(0, X_n)\|_\infty.$$

Since $\|y\| \rightarrow \infty$ implies $\|y\|_\infty \rightarrow \infty$ and $\|F_{\theta,n}(y)\| \geq \|F_{\theta,n}(y)\|_\infty$, it follows that $\|F_{\theta,n}(y)\| \rightarrow \infty$ whenever $\|y\| \rightarrow \infty$. The Hadamard global inverse function theorem therefore implies that $F_{\theta,n}$ is a global C^1 diffeomorphism of \mathbb{R}^n onto itself (Ruzhansky

and Sugimoto 2015, Theorem 2.2).

The derivative formulas (2.1) follow by differentiating $F_{\theta,n}(Y_n^\theta) = \varepsilon_n$ with respect to ε_n and θ and solving for $\partial_\varepsilon Y_n^\theta$ and $\partial_\theta Y_n^\theta$ using the invertible Jacobian $\partial_y F_{\theta,n}(Y_n^\theta) = I_n - A_{\theta,n}(Y_n^\theta, X_n)$.

Proof of Lemma 1''. By Lemma 1', $F_{\theta,n}$ is a C^1 diffeomorphism, and by (B3), ε_n has Lebesgue density $f_{\varepsilon,n}(\varepsilon) = \prod_i f_\varepsilon(\varepsilon_i)$. The change of variables formula applied to the pushforward $Y_n^\theta = F_{\theta,n}^{-1}(\varepsilon_n)$ yields (2.2) (MIT OpenCourseWare 2010, Eq. (7)). Strict positivity follows from $f_\varepsilon > 0$ everywhere together with $\det \partial_y F_{\theta,n}(y) \neq 0$ everywhere. Continuity in (y, θ) follows from continuity of $(y, \theta) \mapsto F_{\theta,n}(y)$, of $(y, \theta) \mapsto A_{\theta,n}(y, X_n)$, and of the determinant map, together with continuity of f_ε . Differentiability in θ follows by the chain rule applied to $f_{\varepsilon,n}(F_{\theta,n}(y))$ and $\det(I_n - A_{\theta,n}(y, X_n))$, using the mixed derivative bound (D1)(iii).

Proof of Lemma 2. Conditioning on (G_n, X_n) , for each focal node u the focal-node-centered graph type $g(u)$ and the label-invariant representation of the covariate restriction are deterministic. The outcome vector Y_n^θ is Borel measurable in ε_n by lemma 1(c), hence for each u the pulled-back attribute vector $a_{n,k}^\theta(u) \in \mathcal{A}_{g(u)}$ is measurable in ε_n . By construction of the quotient σ -algebra, the orbit map $\pi_{g(u)} : \mathcal{A}_{g(u)} \rightarrow \mathcal{A}_{g(u)}/\text{Aut}_\circ(g(u))$ is measurable, so $\varepsilon_n \mapsto S_{n,k}^\theta(u) = (g(u), \pi_{g(u)}(a_{n,k}^\theta(u)))$ is measurable as a map into $(\mathcal{S}_k, \mathcal{F}_k)$. Drawing $U_n \sim \text{Unif}(V_n)$ independently of ε_n , the map $(U_n, \varepsilon_n) \mapsto S_{n,k}^\theta(U_n)$ is measurable as a finite union of measurable graphs indexed by $u \in V_n$. Therefore $P_{n,k}^\theta = \mathcal{L}(S_{n,k}^\theta(U_n) \mid G_n, X_n)$ is a well defined probability measure on $(\mathcal{S}_{n,k}, \mathcal{F}_{n,k}) = (\mathcal{S}_k, \mathcal{F}_k)$, uniquely determined for each θ by the uniqueness of Y_n^θ .

Proof of Lemma 3. Let $\mu := P + Q$. Then $P \ll \mu$ and $Q \ll \mu$. By the Radon-Nikodym theorem (Axler 2020, Theorem 9.36), there exist \mathcal{F} -measurable densities: Set

$$p := \frac{dP}{d\mu}, \quad q := \frac{dQ}{d\mu},$$

so that $p, q \geq 0$ and $p + q = 1$ μ -a.e. For any measurable $D : \mathcal{S} \rightarrow [0, 1]$,

$$V(P, Q; D) = \int_{\mathcal{S}} \left(p(s) \log D(s) + q(s) \log(1 - D(s)) \right) d\mu(s),$$

where terms with $p(s) = 0$ or $q(s) = 0$ are interpreted using $0 \log 0 := 0$.

Fix s with $p(s) + q(s) = 1$ and consider the function

$$g_s(t) := p(s) \log t + q(s) \log(1 - t), \quad t \in [0, 1],$$

with the same boundary conventions. If $p(s) \in (0, 1)$, then g_s is strictly concave on

$(0, 1)$ and its unique maximizer over $[0, 1]$ is the solution to

$$\frac{p(s)}{t} - \frac{q(s)}{1-t} = 0 \iff t = p(s).$$

If $p(s) = 0$, then $g_s(t) = q(s) \log(1-t)$ is maximized at $t = 0 = p(s)$. If $p(s) = 1$, then $g_s(t) = \log t$ is maximized at $t = 1 = p(s)$. Hence in every case and for every $t \in [0, 1]$,

$$g_s(t) \leq g_s(p(s)) = p(s) \log p(s) + q(s) \log q(s).$$

Applying this pointwise bound with $t = D(s)$ and integrating over μ yields, for every measurable $D : \mathcal{S} \rightarrow [0, 1]$,

$$V(P, Q; D) \leq \int_{\mathcal{S}} (p \log p + q \log q) d\mu.$$

Since p is \mathcal{F} -measurable and takes values in $[0, 1]$, the function $D^*(s) := p(s)$ is measurable and admissible. It attains equality μ -a.e. in the pointwise bound, hence attains the supremum. This proves (a) and shows

$$V^*(P, Q) = \int_{\mathcal{S}} (p \log p + q \log q) d\mu.$$

For (b), note that $M = \frac{1}{2}\mu$, so $dP/dM = 2p$ and $dQ/dM = 2q$ M -a.e. Therefore

$$\text{KL}(P \parallel M) = \int \log\left(\frac{dP}{dM}\right) dP = \int p \log(2p) d\mu = \int p \log p d\mu + \log 2,$$

and analogously $\text{KL}(Q \parallel M) = \int q \log q d\mu + \log 2$. Averaging the two expressions gives

$$\text{JS}(P \parallel Q) = \frac{1}{2} \text{KL}(P \parallel M) + \frac{1}{2} \text{KL}(Q \parallel M) = \frac{1}{2} \int (p \log p + q \log q) d\mu + \log 2.$$

Rearranging yields

$$\int (p \log p + q \log q) d\mu = 2 \text{JS}(P \parallel Q) - 2 \log 2 = -\log 4 + 2 \text{JS}(P \parallel Q),$$

which completes the proof.

Proof of Lemma 4. Let $M := \frac{1}{2}(P + Q)$. Then $P \ll M$ and $Q \ll M$. Writing $p := dP/dM$ and $q := dQ/dM$, Jensen's inequality applied to the strictly convex function $t \mapsto t \log t$ yields

$$\int p \log p dM \geq \left(\int p dM \right) \log \left(\int p dM \right) = 0,$$

with equality if and only if p is M -almost surely constant. Since $\int p dM = 1$, equality holds if and only if $p = 1$ M -almost everywhere. This is the information inequality; see Cover and Thomas (2006, Theorem 2.6.3) for the discrete analogue. Therefore

$$\text{KL}(P \parallel M) = \int \log p dP = \int p \log p dM \geq 0,$$

with equality if and only if $P = M$. Applying the same argument to Q yields $\text{KL}(Q \parallel M) \geq 0$, with equality if and only if $Q = M$. Since $\text{JS}(P \parallel Q) = \frac{1}{2}\text{KL}(P \parallel M) + \frac{1}{2}\text{KL}(Q \parallel M)$, we have $\text{JS}(P \parallel Q) = 0$ if and only if $P = M = Q$, that is, if and only if $P = Q$.

Proof of Lemma 5. The mixture representation follows from the definition $P_{n,k}^\theta = \mathcal{L}(S_{n,k}^\theta(U_n) \mid G_n, X_n)$ and the fact that U_n is uniform on V_n and independent of ε_n conditional on (G_n, X_n) . For each focal node u , $\Omega(S_{n,k}^\theta(u)) = \omega_u$ deterministically, so $P_{n,k}^{\theta,u}$ is supported on $\Omega^{-1}(\{\omega_u\})$. The set $\{\omega_u\}$ is measurable in \mathcal{W}_k because it is a singleton in a disjoint union of quotient measurable spaces under finite group actions.

Assume (D3) and fix u . Since $\omega_v \neq \omega_u$ for $v \neq u$, we have $\Omega(S_{n,k}^\theta(v)) \notin \{\omega_u\}$ for all $v \neq u$, and hence

$$\begin{aligned} P_{n,k}^\theta(A \cap \Omega^{-1}(\{\omega_u\})) &= \frac{1}{n} \sum_{v \in V_n} \mathbb{P}_\theta(S_{n,k}^\theta(v) \in A, \Omega(S_{n,k}^\theta(v)) = \omega_u \mid G_n, X_n) \\ &= \frac{1}{n} P_{n,k}^{\theta,u}(A). \end{aligned}$$

If $P_{n,k}^\theta = P_{n,k}^{\theta_0}$, then evaluating the preceding identity at each u and taking A arbitrary yields $P_{n,k}^{\theta,u} = P_{n,k}^{\theta_0,u}$.

Proof of Theorem 1. The argument has two parts. Part I shows that θ_0 achieves the minimum and that equality requires matching distributions of local observations; Part II shows that matching distributions of local observations forces $\theta = \theta_0$.

Part I. By lemma 1, $Y_n^{\theta_0}$ is the unique equilibrium at the true parameter, so the observed distribution of local observations equals $P_{n,k}^{\theta_0}$. By lemma 3(b), $V_{n,k}^*(\theta_0) = -\log 4 + 2 \text{JS}(P_{n,k}^{\theta_0} \parallel P_{n,k}^{\theta_0}) = -\log 4$. Since $\text{JS} \geq 0$, we have $V_{n,k}^*(\theta) \geq -\log 4$ for all θ , with equality if and only if $P_{n,k}^\theta = P_{n,k}^{\theta_0}$ by lemma 4.

Part II. Fix $k \geq k_0$. Let $\pi_{k \rightarrow k_0} : \mathcal{S}_{n,k} \rightarrow \mathcal{S}_{n,k_0}$ denote the measurable restriction map that deletes vertices outside $B_{n,k_0}(u)$ while preserving the focal node. If $P_{n,k}^\theta = P_{n,k}^{\theta_0}$, then the pushforward under $\pi_{k \rightarrow k_0}$ yields $P_{n,k_0}^\theta = P_{n,k_0}^{\theta_0}$. It suffices to prove the contrapositive at radius k_0 . Henceforth set $k := k_0$ and suppress the subscript. We show that for $\theta \in \Theta$ with $0 < \|\theta - \theta_0\| < t_0$, we have $P_{n,k}^\theta \neq P_{n,k}^{\theta_0}$.

Substep (a): Focal node-level distributional shift. Fix $\theta \in \Theta$ with $0 < \|\theta - \theta_0\| < t_0$,

and write $t := \|\theta - \theta_0\|$ and $\delta := (\theta - \theta_0)/\|\theta - \theta_0\|$. By (D2), there exists a focal node $u \in U_\delta$ and a function $g_{u,\delta} \in L^1(\mathbb{R}^{|B_{n,k}(u)|})$ with $\|g_{u,\delta}\|_{L^1} \geq c_g$ such that

$$\|f_{B|\theta_0+t\delta,u} - f_{B|\theta_0,u} - tg_{u,\delta}\|_{L^1} \leq \frac{c_g}{2}t.$$

By the reverse triangle inequality,

$$\|f_{B|\theta_0+t\delta,u} - f_{B|\theta_0,u}\|_{L^1} \geq t\|g_{u,\delta}\|_{L^1} - \|f_{B|\theta_0+t\delta,u} - f_{B|\theta_0,u} - tg_{u,\delta}\|_{L^1} \geq \frac{c_g}{2}t,$$

and therefore

$$\|L_u^{\theta_0+t\delta} - L_u^{\theta_0}\|_{\text{TV}} = \frac{1}{2}\|f_{B|\theta_0+t\delta,u} - f_{B|\theta_0,u}\|_{L^1} \geq \frac{c_g}{4}t > 0.$$

Thus $L_u^\theta = L_u^{\theta_0+t\delta} \neq L_u^{\theta_0}$.

Substep (b): Deconvolution of the mixture. By lemma 5, under (D3) the unit-specific ego-object laws are identified from the mixture distribution of local observations by restricting to the measurable fiber $\Omega^{-1}(\{\omega_u\})$. In particular, $P_{n,k}^\theta = P_{n,k}^{\theta_0}$ implies $P_{n,k}^{\theta,u} = P_{n,k}^{\theta_0,u}$ for every focal node u . Applying the measurable projection that retains only the outcome coordinates on $B_{n,k}(u)$ yields $L_u^\theta = L_u^{\theta_0}$ for every u . Hence $L_u^\theta \neq L_u^{\theta_0}$ for some u implies $P_{n,k}^\theta \neq P_{n,k}^{\theta_0}$.

Conclusion. Fix $\theta \in \Theta$ with $0 < \|\theta - \theta_0\| < t_0$. Substep (a) yields a focal node u with $L_u^\theta \neq L_u^{\theta_0}$. Substep (b) implies that $P_{n,k}^\theta \neq P_{n,k}^{\theta_0}$. Combined with Part I, this establishes that θ_0 is the unique local minimizer of $V_{n,k}^*(\theta)$ on $\{\theta : \|\theta - \theta_0\| < t_0\}$.

Proof of Proposition 1. Fix $\epsilon > 0$ and let $\delta_\infty(\epsilon) > 0$ be as in (A.14). For $n \in \mathbb{N} \cup \{\infty\}$ define $K_n := \{\bar{P}_{n,k}^\theta : \theta \in \Theta\}$, where we interpret $K_\infty := \{\bar{P}_{\infty,k}^\theta : \theta \in \Theta\}$. The Lipschitz bound in the proof of Lemma 12 implies that $\theta \mapsto \bar{P}_{n,k}^\theta$ is continuous in total variation for each fixed n , so each K_n is compact because Θ is compact. Let $\mathcal{K} := \bigcup_{n \in \mathbb{N} \cup \{\infty\}} K_n$. To see that \mathcal{K} is compact, consider a sequence $P_m \in \mathcal{K}$. If the associated indices n_m take only finitely many values, a subsequence lies in a fixed compact set K_n and therefore has a convergent subsequence. Otherwise, pass to a subsequence with $n_m \rightarrow \infty$ and write $P_m = \bar{P}_{n_m,k}^{\theta_m}$. By (A.10), $\|\bar{P}_{n_m,k}^{\theta_m} - \bar{P}_{\infty,k}^{\theta_m}\|_{\text{TV}} \rightarrow 0$, and since K_∞ is compact, $(\bar{P}_{\infty,k}^{\theta_m})_m$ has a convergent subsequence. Along that subsequence, $(P_m)_m$ converges in total variation as well. Thus \mathcal{K} is sequentially compact and hence compact. By Lemma 13, JS is continuous under total variation convergence and therefore uniformly continuous on the compact set $\mathcal{K} \times \mathcal{K}$. Since $\|\bar{P}_{n,k}^{\theta_0} - \bar{P}_{\infty,k}^{\theta_0}\|_{\text{TV}} \rightarrow 0$ and (A.10) holds, uniform continuity yields

$$\sup_{\theta \in \Theta} \left| \text{JS}(\bar{P}_{n,k}^{\theta_0} \parallel \bar{P}_{n,k}^\theta) - \text{JS}(\bar{P}_{\infty,k}^{\theta_0} \parallel \bar{P}_{\infty,k}^\theta) \right| \rightarrow 0.$$

Equivalently,

$$\sup_{\theta \in \Theta} |\bar{V}_{n,k}^*(\theta) - \bar{V}_{\infty,k}^*(\theta)| \longrightarrow 0, \quad \bar{V}_{n,k}^*(\theta) := -\log 4 + 2 \text{JS}(\bar{P}_{n,k}^{\theta_0} \parallel \bar{P}_{n,k}^\theta).$$

Choose $N(\epsilon)$ such that for all $n \geq N(\epsilon)$,

$$\sup_{\theta \in \Theta} |\bar{V}_{n,k}^*(\theta) - \bar{V}_{\infty,k}^*(\theta)| \leq \delta_\infty(\epsilon)/2.$$

Then for every $n \geq N(\epsilon)$ and every θ with $\|\theta - \theta_0\| \geq \epsilon$,

$$\bar{V}_{n,k}^*(\theta) - \bar{V}_{n,k}^*(\theta_0) \geq (\bar{V}_{\infty,k}^*(\theta) - \bar{V}_{\infty,k}^*(\theta_0)) - \delta_\infty(\epsilon) \geq \delta_\infty(\epsilon)/2.$$

Defining $\delta(\epsilon) := \delta_\infty(\epsilon)/2$, this gives the uniform separation bound for the reduced criterion $\bar{V}_{n,k}^*$. Finally, (A.4) implies that for all θ ,

$$V_{n,k}^*(\theta) - V_{n,k}^*(\theta_0) \geq \bar{V}_{n,k}^*(\theta) - \bar{V}_{n,k}^*(\theta_0),$$

so the same $\delta(\epsilon)$ and $N(\epsilon)$ yield the separation inequality in Proposition 1 for the full criterion $V_{n,k}^*$.

Proof of Lemma 6. Part (a) follows from the contraction mapping theorem as in lemma 1. For the bound, the fixed point equation $Y = h_{\theta,n}(Y, X_n) + \varepsilon$ gives $\|Y\|_\infty \leq \|h_{\theta,n}(Y, X_n) - h_{\theta,n}(0, X_n)\|_\infty + \|h_{\theta,n}(0, X_n)\|_\infty + \|\varepsilon\|_\infty \leq \rho\|Y\|_\infty + M_h + \|\varepsilon\|_\infty$, and rearranging yields the claim.

For part (b), subtracting the fixed point equations at θ and θ' with common shocks ε gives $Y_n^\theta - Y_n^{\theta'} = h_{\theta,n}(Y_n^\theta, X_n) - h_{\theta',n}(Y_n^{\theta'}, X_n)$. Adding and subtracting $h_{\theta,n}(Y_n^{\theta'}, X_n)$ and taking ℓ_∞ norms yields $\|Y_n^\theta - Y_n^{\theta'}\|_\infty \leq \rho\|Y_n^\theta - Y_n^{\theta'}\|_\infty + L_\theta\|\theta - \theta'\|$, whence $(1 - \rho)\|Y_n^\theta - Y_n^{\theta'}\|_\infty \leq L_\theta\|\theta - \theta'\|$.

Proof of Lemma 7. Define the Picard iterates $y^{(t)}(\varepsilon)$ by $y^{(0)} = 0$ and $y^{(t+1)} = h_{\theta,n}(y^{(t)}, X_n) + \varepsilon$, and likewise $y^{(t)}(\varepsilon')$. The proof combines two observations.

The first is a finite speed propagation property: for each node i and iterate $t \geq 1$, the value $y_i^{(t)}(\varepsilon)$ depends on ε only through the restriction $\varepsilon|_{B_{n,t-1}(i)}$. This follows by induction on t . At $t = 1$, $y_i^{(1)} = h_{\theta,n,i}(0, X_n) + \varepsilon_i$ depends on ε only through ε_i , the restriction to $B_{n,0}(i)$. If the claim holds at t , then $y_i^{(t+1)} = h_{\theta,n,i}(y^{(t)}, X_n) + \varepsilon_i$. By locality (A2), $h_{\theta,n,i}$ depends on $y^{(t)}$ only through coordinates in $B_{n,1}(i)$. Each such coordinate $y_j^{(t)}$ depends by induction on $\varepsilon|_{B_{n,t-1}(j)}$, and $B_{n,t-1}(j) \subseteq B_{n,t}(i)$ since $d_n(i, j) \leq 1$. Together with $\varepsilon_i \in B_{n,t}(i)$, this establishes the inductive step.

Now fix $i \in B_{n,k}(u)$. Since $d_n(i, u) \leq k$, we have $B_{n,r}(i) \subseteq B_{n,k+r}(u)$, so the shocks ε and ε' agree on $B_{n,r}(i)$. By the finite speed property at $t = r + 1$, the iterate $y_i^{(r+1)}(\varepsilon)$ depends only on $\varepsilon|_{B_{n,r}(i)}$, and hence $y_i^{(r+1)}(\varepsilon) = y_i^{(r+1)}(\varepsilon')$.

The second observation is a contraction tail bound. Since $\Gamma_{\theta,n}$ is a contraction with modulus ρ in ℓ_∞ , successive iterates satisfy $\|y^{(t+1)} - y^{(t)}\|_\infty \leq \rho^t\|y^{(1)} - y^{(0)}\|_\infty$.

Summing the geometric series from $t = r + 1$ onward gives

$$\|Y_n^\theta(\varepsilon) - y^{(r+1)}(\varepsilon)\|_\infty \leq \frac{\rho^{r+1}}{1-\rho} \|y^{(1)}(\varepsilon)\|_\infty = \frac{\rho^{r+1}}{1-\rho} (M_h + \|\varepsilon\|_\infty),$$

where the last equality uses $\|y^{(1)}(\varepsilon)\|_\infty = \|h_{\theta,n}(0, X_n) + \varepsilon\|_\infty \leq M_h + \|\varepsilon\|_\infty$ from (C2).

Combining the two parts by the triangle inequality, for each $i \in B_{n,k}(u)$,

$$|Y_i^\theta(\varepsilon) - Y_i^\theta(\varepsilon')| \leq |Y_i^\theta(\varepsilon) - y_i^{(r+1)}(\varepsilon)| + \underbrace{|y_i^{(r+1)}(\varepsilon) - y_i^{(r+1)}(\varepsilon')|}_{=0} + |y_i^{(r+1)}(\varepsilon') - Y_i^\theta(\varepsilon')|,$$

and the two nonzero terms are each bounded by $\frac{\rho^{r+1}}{1-\rho} (M_h + \|\varepsilon\|_\infty)$ and $\frac{\rho^{r+1}}{1-\rho} (M_h + \|\varepsilon'\|_\infty)$ respectively. Taking the maximum over $i \in B_{n,k}(u)$ yields (3.1).

Proof of Lemma 8. By the clipping condition $D_\phi \in [\eta, 1 - \eta]$, the maps $s \mapsto \log D_\phi(s)$ and $s \mapsto \log(1 - D_\phi(s))$ are Lipschitz with constant L_D/η on $(\mathcal{S}_{n,k}, d_S)$. By the definition of d_S in (C5), for any $u \in V_n$ and any outcome vectors $y, y' \in \mathbb{R}^n$,

$$|\log D_\phi(S_{n,k}(u; y)) - \log D_\phi(S_{n,k}(u; y'))| \leq \frac{L_D}{\eta} \|y - y'\|_{\infty, B_{n,k}(u)},$$

and the same bound holds with $\log(1 - D_\phi)$ in place of $\log D_\phi$. Applying lemma 7 separately to the observed shocks $(\varepsilon, \varepsilon^{(u,r)})$ and the simulated shocks $(\tilde{\varepsilon}, \tilde{\varepsilon}^{(u,r)})$, which by construction agree on $B_{n,k+r}(u)$, and combining the two Lipschitz bounds yields (3.2) with $C_1 = 2C_0L_D/\eta$.

Proof of Lemma 9. For part (a), if $B_{n,k+r}(u) \cap B_{n,k+r}(v) \neq \emptyset$, there exists w with $d_n(u, w) \leq k + r$ and $d_n(v, w) \leq k + r$, so $d_n(u, v) \leq 2(k + r)$. Every neighbor of u in $H_{n,r}$ therefore lies in $B_{n,2(k+r)}(u) \setminus \{u\}$, giving $\deg_{H_{n,r}}(u) \leq |B_{n,2(k+r)}(u)| - 1 \leq \nu_n(2(k + r)) - 1$. To construct blocks, apply a greedy coloring of $H_{n,r}$. At each step, at most $\Delta_{n,r}$ labels are excluded by already assigned neighbors, so one label among $\Delta_{n,r} + 1$ is available. Therefore $\chi(H_{n,r}) \leq \Delta_{n,r} + 1 \leq \nu_n(2(k + r))$.

For part (b), fix a greedy blocking partition into $q_{n,r}$ blocks. Within any block $C_{n,r,j}$, the balls $\{B_{n,k+r}(u) : u \in C_{n,r,j}\}$ are pairwise disjoint by construction. By the local approximation with influence decay, the loss $\ell_{\theta,\phi}^{(u,r)}(u)$ is measurable with respect to the shocks $\varepsilon|_{B_{n,k+r}(u)}$ together with an independent outside copy $\varepsilon^{(u,\text{out})}$ specific to focal node u . Since (C4) posits that shocks are i.i.d. conditional on (G_n, X_n) , disjointness of the shock supports implies that the truncated losses within each block are conditionally independent.

Proof of Lemma 10. The proof has four steps.

Since $D_\phi \in [\eta, 1 - \eta]$, each summand $\ell_{\theta,\phi}^{(u,r)}(u)$ is bounded in $[2 \log \eta, 2 \log(1 - \eta)]$, uniformly in θ, ϕ, u , and n . In particular, $|\ell_{\theta,\phi}^{(u,r)}(u)| \leq B$ with $B := 2|\log \eta|$.

Fix (θ, ϕ) and a truncation depth r . Define the centered variables

$$Z_u := \ell_{\theta, \phi}^{(u, r)}(u) - \mathbb{E}[\ell_{\theta, \phi}^{(u, r)}(u) \mid G_n, X_n], \quad u \in V_n.$$

Then $\mathbb{E}[Z_u \mid G_n, X_n] = 0$ and $Z_u \in [-2B, 2B]$ because $|Z_u| \leq |\ell_{\theta, \phi}^{(u, r)}(u)| + \mathbb{E}[|\ell_{\theta, \phi}^{(u, r)}(u)| \mid G_n, X_n] \leq 2B$. Fix $\lambda > 0$. By convexity of the exponential function on $[-2B, 2B]$, for each $z \in [-2B, 2B]$,

$$e^{\lambda z} \leq \frac{2B - z}{4B} e^{-2\lambda B} + \frac{2B + z}{4B} e^{2\lambda B} = \cosh(2\lambda B) + \frac{z}{2B} \sinh(2\lambda B).$$

Taking conditional expectations and using $\mathbb{E}[Z_u \mid G_n, X_n] = 0$ yields $\mathbb{E}[e^{\lambda Z_u} \mid G_n, X_n] \leq \cosh(2\lambda B)$. Since $\cosh(x) = \sum_{m \geq 0} x^{2m} / (2m)! \leq \sum_{m \geq 0} (x^2/2)^m / m! = e^{x^2/2}$, we obtain $\cosh(2\lambda B) \leq e^{2\lambda^2 B^2}$ and hence $\mathbb{E}[e^{\lambda Z_u} \mid G_n, X_n] \leq e^{2\lambda^2 B^2}$. Define the blocking partition $V_n = \bigcup_{j=1}^{q_{n,r}} C_{n,r,j}$ from lemma 9(b). Conditional on (G_n, X_n) , the collection $\{Z_u : u \in C_{n,r,j}\}$ is independent for each j . Since $\exp(\lambda \sum_{u \in V_n} Z_u) = \prod_{j=1}^{q_{n,r}} \exp(\lambda \sum_{u \in C_{n,r,j}} Z_u)$, Hölder's inequality with exponents $(q_{n,r}, \dots, q_{n,r})$ yields

$$\begin{aligned} \mathbb{E}\left[e^{\lambda \sum_{u \in V_n} Z_u} \mid G_n, X_n\right] &\leq \prod_{j=1}^{q_{n,r}} \mathbb{E}\left[e^{q_{n,r} \lambda \sum_{u \in C_{n,r,j}} Z_u} \mid G_n, X_n\right]^{1/q_{n,r}} \\ &\leq \prod_{j=1}^{q_{n,r}} \exp(2q_{n,r} \lambda^2 B^2 |C_{n,r,j}|) = \exp(2q_{n,r} \lambda^2 B^2 n), \end{aligned}$$

where the second inequality uses conditional independence within $C_{n,r,j}$ and $\mathbb{E}[e^{q_{n,r} \lambda Z_u} \mid G_n, X_n] \leq e^{2q_{n,r}^2 \lambda^2 B^2}$. Therefore, by Markov's inequality applied to $\exp(\lambda \sum_{u \in V_n} Z_u)$ (Vershynin 2025, Proposition 1.6.2),

$$\begin{aligned} \mathbb{P}\left(\widehat{V}_{n,k}^{(r)}(\theta, \phi) - V_{n,k}^{(r)}(\theta, \phi) > t \mid G_n, X_n\right) &= \mathbb{P}\left(\sum_{u \in V_n} Z_u > nt \mid G_n, X_n\right) \\ &\leq \exp(-\lambda nt) \mathbb{E}\left[e^{\lambda \sum_{u \in V_n} Z_u} \mid G_n, X_n\right] \\ &\leq \exp(-\lambda nt + 2q_{n,r} \lambda^2 B^2 n). \end{aligned}$$

Optimizing at $\lambda = t/(4q_{n,r} B^2)$ yields

$$\mathbb{P}\left(\widehat{V}_{n,k}^{(r)}(\theta, \phi) - V_{n,k}^{(r)}(\theta, \phi) > t \mid G_n, X_n\right) \leq \exp(-c_\eta m_{\text{eff},n}(r) t^2),$$

where $c_\eta := 1/(8B^2)$ and $m_{\text{eff},n}(r) = n/q_{n,r}$ as in (3.3). Applying the same bound to $-Z_u$ gives the two sided inequality

$$\mathbb{P}\left(|\widehat{V}_{n,k}^{(r)}(\theta, \phi) - V_{n,k}^{(r)}(\theta, \phi)| > t \mid G_n, X_n\right) \leq 2 \exp(-c_\eta m_{\text{eff},n}(r) t^2).$$

To extend pointwise concentration to uniform convergence over (θ, ϕ) at truncation depth r_n , we discretize both Θ and the adaptive test function induced log function class. Fix $\epsilon > 0$. Let Θ_{ϵ/L_n} be an (ϵ/L_n) -net of Θ under $\|\cdot\|$ with cardinality $|\Theta_{\epsilon/L_n}| = N(\epsilon/L_n, \Theta, \|\cdot\|)$. Let $\mathcal{F}_{n,\epsilon} \subseteq \mathcal{F}_n$ be an ϵ -net of \mathcal{F}_n under $\|\cdot\|_\infty$ with cardinality $|\mathcal{F}_{n,\epsilon}| = N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty)$. For each focal node u and depth r , let $S_{n,k}^{\theta_0, (u,r)}(u)$ and $\tilde{S}_{n,k}^{\theta, (u,r)}(u)$ denote the truncated observed and simulated ego objects from lemma 8, so that

$$\ell_{\theta,\phi}^{(u,r)}(u) = \log D_\phi(S_{n,k}^{\theta_0, (u,r)}(u)) + \log(1 - D_\phi(\tilde{S}_{n,k}^{\theta, (u,r)}(u))).$$

For $f, g : \mathcal{S}_{n,k} \rightarrow \mathbb{R}$, define the corresponding truncated loss

$$\ell_{\theta,f,g}^{(u,r)}(u) := f(S_{n,k}^{\theta_0, (u,r)}(u)) + g(\tilde{S}_{n,k}^{\theta, (u,r)}(u)).$$

Define the induced criteria by

$$\widehat{V}_{n,k}^{(r)}(\theta, f, g) := \frac{1}{n} \sum_{u \in V_n} \ell_{\theta,f,g}^{(u,r)}(u), \quad V_{n,k}^{(r)}(\theta, f, g) := \mathbb{E}[\widehat{V}_{n,k}^{(r)}(\theta, f, g) \mid G_n, X_n],$$

The two sided concentration inequality derived above applies to $\widehat{V}_{n,k}^{(r)}(\theta, f, g) - V_{n,k}^{(r)}(\theta, f, g)$ verbatim because it uses only boundedness and the blocking partition with within-block conditional independence from lemma 9(b).

For each $\phi \in \Phi_n$, choose $f_{\phi,\epsilon}, g_{\phi,\epsilon} \in \mathcal{F}_{n,\epsilon}$ such that $\|f_{\phi,\epsilon} - \log D_\phi\|_\infty \leq \epsilon$ and $\|g_{\phi,\epsilon} - \log(1 - D_\phi)\|_\infty \leq \epsilon$. Then $\sup_u |\ell_{\theta,\phi}^{(u,r)}(u) - \ell_{\theta,f_{\phi,\epsilon},g_{\phi,\epsilon}}^{(u,r)}(u)| \leq 2\epsilon$, and hence

$$|(\widehat{V}_{n,k}^{(r)} - V_{n,k}^{(r)})(\theta, \phi) - (\widehat{V}_{n,k}^{(r)} - V_{n,k}^{(r)})(\theta, f_{\phi,\epsilon}, g_{\phi,\epsilon})| \leq 4\epsilon.$$

Similarly, for any θ and any $\theta_{\epsilon/L_n} \in \Theta_{\epsilon/L_n}$ with $\|\theta - \theta_{\epsilon/L_n}\| \leq \epsilon/L_n$, on the event \mathcal{A}_n from (C3') we have $\sup_{u,\phi} |\ell_{\theta,\phi}^{(u,r_n)}(u) - \ell_{\theta_{\epsilon/L_n},\phi}^{(u,r_n)}(u)| \leq \epsilon$, and hence

$$|(\widehat{V}_{n,k}^{(r_n)} - V_{n,k}^{(r_n)})(\theta, \phi) - (\widehat{V}_{n,k}^{(r_n)} - V_{n,k}^{(r_n)})(\theta_{\epsilon/L_n}, \phi)| \leq 2\epsilon.$$

On \mathcal{A}_n , the two discretization bounds combine to an off net control by 6ϵ . A union bound (Vershynin 2025, Lemma 1.4.1) over Θ_{ϵ/L_n} and over pairs $(f, g) \in \mathcal{F}_{n,\epsilon} \times \mathcal{F}_{n,\epsilon}$ therefore yields

$$\begin{aligned} \mathbb{P}\left(\sup_{\theta,\phi} |\widehat{V}_{n,k}^{(r_n)}(\theta, \phi) - V_{n,k}^{(r_n)}(\theta, \phi)| > t + 6\epsilon \mid G_n, X_n\right) \\ \leq \mathbb{P}(\mathcal{A}_n^c \mid G_n, X_n) \\ + 2N(\epsilon/L_n, \Theta, \|\cdot\|)N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty)^2 \exp(-c_\eta m_{\text{eff},n}(r_n)t^2). \end{aligned}$$

Fix $\delta > 0$ and apply the bound with $t := \delta/2$ and $\epsilon := \delta/12$, so that $t + 6\epsilon = \delta$. The entropy condition (3.4) implies that the logarithm of the prefactor is $o(m_{\text{eff},n}(r_n))$, while

the exponential term equals $\exp(-c_\eta m_{\text{eff},n}(r_n)\delta^2/4)$. In addition, $\mathbb{P}(\mathcal{A}_n^c \mid G_n, X_n) \rightarrow 0$ by (C3'). Therefore the right side converges to zero, so $\sup_{\theta,\phi} |\widehat{V}_{n,k}^{(r_n)}(\theta, \phi) - V_{n,k}^{(r_n)}(\theta, \phi)| \xrightarrow{P} 0$.

Proof of Lemma 11. From the truncation error bound (3.2) in lemma 8, the per focal node bound holds with unit-specific outside copies $\varepsilon^{(u,\text{out})}$ and $\tilde{\varepsilon}^{(u,\text{out})}$ as introduced in that lemma. The empirical truncation gap therefore satisfies

$$\begin{aligned} \sup_{\theta,\phi} |\widehat{V}_{n,k}(\theta, \phi) - \widehat{V}_{n,k}^{(r_n)}(\theta, \phi)| \\ \leq \frac{1}{n} \sum_{u \in V_n} C_1 \rho^{r_n+1} \left(1 + \|\varepsilon\|_\infty + \|\varepsilon^{(u,\text{out})}\|_\infty \right. \\ \left. + \|\tilde{\varepsilon}\|_\infty + \|\tilde{\varepsilon}^{(u,\text{out})}\|_\infty \right). \end{aligned}$$

Each outside copy $\varepsilon^{(u,\text{out})}$ has the same marginal distribution as ε_n and is independent across focal nodes (by construction), so $\frac{1}{n} \sum_u \|\varepsilon^{(u,\text{out})}\|_\infty$ has expectation $\mathbb{E}[\|\varepsilon_n\|_\infty \mid G_n, X_n]$, and likewise for $\tilde{\varepsilon}^{(u,\text{out})}$. The shock tail condition (C4) gives $\rho^{r_n} \|\varepsilon\|_\infty \xrightarrow{P} 0$ and $\rho^{r_n} \mathbb{E}[\|\varepsilon_n\|_\infty \mid G_n, X_n] \rightarrow 0$, and the same holds for each remaining term, so the right side converges to zero in probability. The population version follows by the same bound applied under the conditional expectation, using $\rho^{r_n} \mathbb{E}[\|\varepsilon_n\|_\infty \mid G_n, X_n] \rightarrow 0$ from (C4).

Proof of Corollary 2. For each θ , $|\widehat{V}_{n,k}^{\Phi_n}(\theta) - V_{n,k}^{\Phi_n}(\theta)| = |\sup_\phi \widehat{V}_{n,k}(\theta, \phi) - \sup_\phi V_{n,k}(\theta, \phi)| \leq \sup_\phi |\widehat{V}_{n,k}(\theta, \phi) - V_{n,k}(\theta, \phi)|$. Taking \sup_θ and bounding by $\sup_{\theta,\phi} |\widehat{V}_{n,k} - V_{n,k}|$, the claim follows from the triangle inequality through the truncated intermediate:

$$\sup_{\theta,\phi} |\widehat{V}_{n,k} - V_{n,k}| \leq \sup_{\theta,\phi} |\widehat{V}_{n,k} - \widehat{V}_{n,k}^{(r_n)}| + \sup_{\theta,\phi} |\widehat{V}_{n,k}^{(r_n)} - V_{n,k}^{(r_n)}| + \sup_{\theta,\phi} |V_{n,k}^{(r_n)} - V_{n,k}|,$$

each term of which vanishes by lemma 11, lemma 10, and lemma 11 respectively.

Proof of Theorem 2. Fix $\epsilon > 0$ and let $\delta := \delta(\epsilon) > 0$ and $N_{\text{sep}} := N(\epsilon)$ be as in Proposition 1. Define the event

$$\mathcal{E}_n := \left\{ \sup_{\theta \in \Theta} |\widehat{V}_{n,k}^{\Phi_n}(\theta) - V_{n,k}^{\Phi_n}(\theta)| \leq \delta/4 \right\} \cap \{ \alpha_n \leq \delta/4 \} \cap \{ \xi_n \leq \delta/8 \}.$$

By corollary 2, the first event has probability tending to one. By (C7), $\alpha_n \rightarrow 0$, so there exists $N_\alpha \in \mathbb{N}$ such that $\alpha_n \leq \delta/4$ for all $n \geq N_\alpha$. By the approximate argmin condition, $\xi_n \xrightarrow{P} 0$, so $\mathbb{P}(\xi_n \leq \delta/8) \rightarrow 1$ and hence $\mathbb{P}(\mathcal{E}_n) \rightarrow 1$.

Fix $n \geq N_{\text{sep}}$. On \mathcal{E}_n , take any $\theta \in \Theta$ with $\|\theta - \theta_0\| \geq \epsilon$. We chain four inequalities. First, the definition of the sieve approximation error gives $V_{n,k}^{\Phi_n}(\theta) \geq V_{n,k}^*(\theta) - \alpha_n$, since $V_{n,k}^*(\theta)$ is the supremum over all measurable adaptive test functions while $V_{n,k}^{\Phi_n}(\theta)$

restricts to Φ_n , and $|V_{n,k}^* - V_{n,k}^{\Phi_n}| \leq \alpha_n$ by definition. Second, uniform separation in Proposition 1 gives $V_{n,k}^*(\theta) \geq V_{n,k}^*(\theta_0) + \delta$. Third, $V_{n,k}^*(\theta_0) \geq V_{n,k}^{\Phi_n}(\theta_0)$ because the infinite capacity criterion dominates the sieve restricted one. Assembling:

$$V_{n,k}^{\Phi_n}(\theta) \geq V_{n,k}^*(\theta_0) + \delta - \alpha_n \geq V_{n,k}^{\Phi_n}(\theta_0) + \delta - \alpha_n \geq V_{n,k}^{\Phi_n}(\theta_0) + \frac{3}{4}\delta, \quad (\text{A.2})$$

where the last step in (A.2) uses $\alpha_n \leq \delta/4$ on \mathcal{E}_n . Converting from population to empirical by the uniform bound on \mathcal{E}_n :

$$\widehat{V}_{n,k}^{\Phi_n}(\theta) \geq V_{n,k}^{\Phi_n}(\theta) - \delta/4 \geq V_{n,k}^{\Phi_n}(\theta_0) + \delta/2 \geq \widehat{V}_{n,k}^{\Phi_n}(\theta_0) - \delta/4 + \delta/2 = \widehat{V}_{n,k}^{\Phi_n}(\theta_0) + \delta/4. \quad (\text{A.3})$$

Since (A.3) holds for every θ with $\|\theta - \theta_0\| \geq \epsilon$, if $\|\hat{\theta}_{n,k} - \theta_0\| \geq \epsilon$ then applying (A.3) at $\theta = \hat{\theta}_{n,k}$ gives $\widehat{V}_{n,k}^{\Phi_n}(\hat{\theta}_{n,k}) \geq \widehat{V}_{n,k}^{\Phi_n}(\theta_0) + \delta/4$. On the other hand, because $\theta_0 \in \Theta$ is a candidate in the infimum, the approximate argmin condition implies $\widehat{V}_{n,k}^{\Phi_n}(\hat{\theta}_{n,k}) \leq \widehat{V}_{n,k}^{\Phi_n}(\theta_0) + \xi_n \leq \widehat{V}_{n,k}^{\Phi_n}(\theta_0) + \delta/8$ on \mathcal{E}_n , which contradicts $\delta/8 < \delta/4$. Therefore $\|\hat{\theta}_{n,k} - \theta_0\| < \epsilon$ on \mathcal{E}_n , and $\mathbb{P}(\|\hat{\theta}_{n,k} - \theta_0\| \geq \epsilon) \leq 1 - \mathbb{P}(\mathcal{E}_n) \rightarrow 0$, establishing $\hat{\theta}_{n,k} \xrightarrow{P} \theta_0$. This is the standard extremum-estimator consistency argument; see Newey and McFadden (1994, Theorem 2.1).

Uniform separation via a limiting distribution of local observations.

This appendix proves Proposition 1. Fix an ego radius $k \geq 1$ throughout. The proof proceeds by passing to a limiting law for a reduced k -local object, showing uniform convergence of the associated concentrated population minimum-distance criterion, and transferring strict identification of the limiting criterion back to finite n .

A reduced distribution of local observations.

The concentrated population minimum-distance criterion $V_{n,k}^*(\theta)$ is defined on the full ego object space $\mathcal{S}_{n,k}$, which includes the focal-node-centered subgraph, covariates, and outcomes. For the separation inequality it is convenient to work with a reduced object that drops the covariates while retaining the focal-node-centered k -ball graph and the outcome subvector. Under (A1), there are finitely many focal-node-centered k -ball graph label-invariant structural types. Let \mathcal{G}_k denote this finite set, and for $g \in \mathcal{G}_k$ write $m(g) := |V(g)|$ and let $\text{Aut}_o(g)$ denote the finite group of focal-node-preserving within-neighborhood relabeling symmetries of g . Define the outcome fiber $\bar{\mathcal{A}}_g := \mathbb{R}^{m(g)}$ equipped with the Borel σ -algebra, and let $\text{Aut}_o(g)$ act on $\bar{\mathcal{A}}_g$ by permuting coordinates. Let $\pi_{Y,g} : \bar{\mathcal{A}}_g \rightarrow \bar{\mathcal{A}}_g/\text{Aut}_o(g)$ be the orbit map and equip the quotient with the quotient σ -algebra. Define the reduced ego object space

$$\bar{\mathcal{S}}_k := \bigsqcup_{g \in \mathcal{G}_k} (\{g\} \times \bar{\mathcal{A}}_g/\text{Aut}_o(g))$$

equipped with the disjoint union σ -algebra. For each $g \in \mathcal{G}_k$, let $\text{Leb}_{m(g)}$ denote Lebesgue measure on $\mathbb{R}^{m(g)}$ and set $\lambda_g := (\pi_{Y,g})_{\#} \text{Leb}_{m(g)}$. Define λ on $\bar{\mathcal{S}}_k$ as counting

measure on \mathcal{G}_k times λ_g on each fiber.

For each focal node $u \in V_n$, let $g(u) \in \mathcal{G}_k$ denote the focal-node-centered label-invariant structural type of $G_n[B_{n,k}(u)]$. Choose any focal-node-preserving isomorphism $\iota_u : g(u) \rightarrow G_n[B_{n,k}(u)]$ and pull back the outcome restriction along ι_u to obtain a vector $y_{n,k}^\theta(u) \in \bar{\mathcal{A}}_{g(u)}$. Define

$$\bar{S}_{n,k}^\theta(u) := (g(u), \pi_{Y,g(u)}(y_{n,k}^\theta(u))) \in \bar{\mathcal{S}}_k,$$

which does not depend on the choice of ι_u . Let $\bar{P}_{n,k}^\theta$ denote the law of $\bar{S}_{n,k}^\theta(u)$ under a uniform focal node draw, conditional on (G_n, X_n) . Let $\pi : \mathcal{S}_{n,k} \rightarrow \bar{\mathcal{S}}_k$ be the measurable projection that drops the covariate coordinates, so that $\pi(S_{n,k}^\theta(u)) = \bar{S}_{n,k}^\theta(u)$ and $\pi_\# P_{n,k}^\theta = \bar{P}_{n,k}^\theta$. By Lemma 3, applied to the pushforward measures and using that $V^*(P, Q)$ is a supremum over adaptive test functions on the ambient space, we have for every $\theta \in \Theta$,

$$V_{n,k}^*(\theta) = V^*(P_{n,k}^{\theta_0}, P_{n,k}^\theta) \geq V^*(\bar{P}_{n,k}^{\theta_0}, \bar{P}_{n,k}^\theta) = -\log 4 + 2 \text{JS}(\bar{P}_{n,k}^{\theta_0} \parallel \bar{P}_{n,k}^\theta). \quad (\text{A.4})$$

Since $V_{n,k}^*(\theta_0) = -\log 4$, the separation inequality in Proposition 1 follows once we establish an analogous separation bound for $\text{JS}(\bar{P}_{n,k}^{\theta_0} \parallel \bar{P}_{n,k}^\theta)$.

Additional limit conditions.

We impose three conditions that deliver a limiting reduced distribution of local observations and a strictly identified limiting criterion. For $r \geq 0$ define the Picard iterates $y^{(t)}$ by $y^{(0)} := 0$ and $y^{(t+1)} := h_{\theta,n}(y^{(t)}, X_n) + \varepsilon_n$, and set $Y_n^{\theta,(r)} := y^{(r+1)}$. For each focal node $u \in V_n$ define the truncated reduced ego object

$$\bar{S}_{n,k}^{\theta,(r)}(u) := (g(u), \pi_{Y,g(u)}(y_{n,k}^{\theta,(r)}(u))),$$

where $y_{n,k}^{\theta,(r)}(u) \in \bar{\mathcal{A}}_{g(u)}$ is obtained by pulling back the outcome restriction $(Y_{n,j}^{\theta,(r)})_{j \in B_{n,k}(u)}$ along any focal-node-preserving isomorphism $\iota_u : g(u) \rightarrow G_n[B_{n,k}(u)]$. Let $\bar{P}_{n,k}^{\theta,(r)}$ denote its law under a uniform focal node draw, conditional on (G_n, X_n) .

(P1) Local weak convergence (projective family).

For each radius $r \geq 0$ and focal node $u \in V_n$, let $\omega_{n,u}^{(r)}$ denote the marked distance- r ego neighborhood centered at u , consisting of the induced subgraph $G_n[B_{n,r}(u)]$ together with the covariate marks $(X_{n,j})_{j \in B_{n,r}(u)}$, viewed up to focal-node-preserving isomorphism. Assume there exists a projectively consistent family of probability measures $(\mu_r)_{r \geq 0}$ such that for every r and every bounded continuous test function φ on the marked distance- r ego-neighborhood space,

$$\frac{1}{n} \sum_{u \in V_n} \varphi(\omega_{n,u}^{(r)}) \longrightarrow \int \varphi(\omega) d\mu_r(\omega). \quad (\text{A.5})$$

In addition, assume that for each (θ, r) and λ -a.e. $s \in \bar{\mathcal{S}}_k$ the conditional den-

sity $q_{\theta,r}(s \mid \omega)$ of $\bar{S}_{n,k}^{\theta,(r)}(u)$ given $\omega_{n,u}^{(k+r)} = \omega$ exists with respect to λ , and that $\omega \mapsto q_{\theta,r}(s \mid \omega)$ is bounded and continuous on the support of μ_{k+r} . For the linear Gaussian specialization in Section 5, the conditional law of the labeled outcome vector $y_{n,k}^{\theta,(r)}(u) \in \mathbb{R}^{m(g)}$ given $\omega_{n,u}^{(k+r)} = \omega$ is Gaussian with mean and covariance depending continuously on ω . The induced λ -density $q_{\theta,r}(\cdot \mid \omega)$ on \bar{S}_k is then the corresponding $\text{Aut}_\circ(g)$ -orbit average, so verifying the boundedness and continuity clause reduces to bounding these Gaussian parameter maps on the support of μ_{k+r} .

(P2) Uniform envelopes and total-variation truncation for k -ego outcome densities.

For each n and focal node u , write $B := B_{n,k}(u)$ and let $L_u^\theta := \mathcal{L}((Y_{n,j}^\theta)_{j \in B} \mid G_n, X_n)$ denote the unit-specific k -ego outcome law. Under (B3) and (D1), L_u^θ admits a Lebesgue density $f_{B|\theta,u}$ by Corollary 1. For each $r \geq 0$, let $L_u^{\theta,(r)} := \mathcal{L}((Y_{n,j}^{\theta,(r)})_{j \in B} \mid G_n, X_n)$ denote the truncated counterpart, and assume that $L_u^{\theta,(r)}$ admits a Lebesgue density $f_{B|\theta,u}^{(r)}$. Write $d_k := \sup_n \nu_n(k) < \infty$, which is finite by (A1) for fixed k . Assume there exist envelopes $F_m \in L^1(\mathbb{R}^m)$ for $m = 1, \dots, d_k$ and a constant $\tilde{C} < \infty$ such that for every n, u, θ and every $r \geq 0$ with $|B| = m$,

$$\begin{aligned} \max\{f_{B|\theta,u}(y), f_{B|\theta,u}^{(r)}(y)\} &\leq F_m(y), \quad \text{and} \\ \|\partial_\theta f_{B|\theta,u}(y)\| &\leq \tilde{C} F_m(y) \quad \text{for Lebesgue-a.e. } y \in \mathbb{R}^m. \end{aligned} \quad (\text{A.6})$$

Assume there exists a constant $C_{\text{TV}} < \infty$ such that for all $r \geq 0$,

$$\sup_n \sup_{u \in V_n} \sup_{\theta \in \Theta} \|L_u^\theta - L_u^{\theta,(r)}\|_{\text{TV}} \leq C_{\text{TV}} \rho^{r+1}. \quad (\text{A.7})$$

For the linear Gaussian specialization in Section 5, (A.6) holds with Gaussian envelopes that are uniform over (n, u, θ) on compact Θ , and (A.7) can be verified by bounding total variation between Gaussian laws whose mean and covariance differ by $O(\rho^{r+1})$.

(P3) Limiting identification for the reduced k -distribution of local observations.

Let $\bar{P}_{\infty,k}^\theta$ denote the limiting reduced distribution of local observations constructed in Lemma 12 below. Assume that

$$\bar{P}_{\infty,k}^\theta = \bar{P}_{\infty,k}^{\theta_0} \implies \theta = \theta_0. \quad (\text{A.8})$$

Limiting reduced distributions of local observations.

The separation transfer argument uses continuity of Jensen–Shannon divergence, which we obtain from total variation convergence of the reduced distributions of local observations. The next lemma constructs the limiting reduced distributions of local observations by truncation and establishes convergence in total variation, including uniformity over compact Θ .

Lemma 12 (Limiting reduced distribution of local observations and total variation convergence). *Assume (A1), (A2), (B1), (B2), (B3), (D1), (P1), and (P2). Then for each fixed $\theta \in \Theta$ there exists a unique probability measure $\bar{P}_{\infty,k}^\theta$ such that*

$$\|\bar{P}_{n,k}^\theta - \bar{P}_{\infty,k}^\theta\|_{\text{TV}} \longrightarrow 0. \quad (\text{A.9})$$

If in addition Θ is compact as in (C1), then

$$\sup_{\theta \in \Theta} \|\bar{P}_{n,k}^\theta - \bar{P}_{\infty,k}^\theta\|_{\text{TV}} \longrightarrow 0. \quad (\text{A.10})$$

Proof of Lemma 12. Fix θ . For $r \geq 0$, recall the truncated reduced object

$$\bar{S}_{n,k}^{\theta,(r)}(u) := (g(u), \pi_{Y,g(u)}(y_{n,k}^{\theta,(r)}(u))) \in \bar{\mathcal{S}}_k$$

and its law $\bar{P}_{n,k}^{\theta,(r)} := \frac{1}{n} \sum_{u \in V_n} \mathcal{L}(\bar{S}_{n,k}^{\theta,(r)}(u) \mid G_n, X_n)$. By the finite speed propagation argument in the proof of Lemma 7 (which uses (A2)), the vector $(Y_{n,j}^{\theta,(r)})_{j \in B_{n,k}(u)}$ depends on (G_n, X_n) only through $\omega_{n,u}^{(k+r)}$ and on shocks only through $\varepsilon_{B_{n,k+r}(u)}$. Consequently, there exists a conditional law $\Pi_{\theta,r}(\omega)$ of $\bar{S}_{n,k}^{\theta,(r)}(u)$ given $\omega_{n,u}^{(k+r)} = \omega$ that depends only on ω . By (P1), $\Pi_{\theta,r}(\omega)$ admits a λ -density $q_{\theta,r}(\cdot \mid \omega)$, and for λ -a.e. s the map $\omega \mapsto q_{\theta,r}(s \mid \omega)$ is bounded and continuous on the support of μ_{k+r} . Since $\bar{P}_{n,k}^{\theta,(r)} = \frac{1}{n} \sum_{u \in V_n} \Pi_{\theta,r}(\omega_{n,u}^{(k+r)})$, it admits the mixture density

$$\bar{p}_{n,\theta}^{(r)}(s) = \frac{1}{n} \sum_{u \in V_n} q_{\theta,r}(s \mid \omega_{n,u}^{(k+r)}), \quad s \in \bar{\mathcal{S}}_k.$$

Define $\bar{p}_{\infty,\theta}^{(r)}(s) := \int q_{\theta,r}(s \mid \omega) d\mu_{k+r}(\omega)$ and let $\bar{P}_{\infty,k}^{\theta,(r)}$ be the corresponding probability law on $\bar{\mathcal{S}}_k$. For λ -a.e. s , applying (A.5) with test function $\varphi(\omega) := q_{\theta,r}(s \mid \omega)$ gives $\bar{p}_{n,\theta}^{(r)}(s) \rightarrow \bar{p}_{\infty,\theta}^{(r)}(s)$. Define the envelope \bar{F} on $\bar{\mathcal{S}}_k$ by setting, for $s = (g, \pi_{Y,g}(y))$,

$$\bar{F}(s) := \frac{1}{|\text{Aut}_o(g)|} \sum_{\tau \in \text{Aut}_o(g)} F_{m(g)}(\tau y).$$

This is well defined and integrable with respect to λ because \mathcal{G}_k is finite, each $\text{Aut}_o(g)$ is finite, and $F_m \in L^1(\mathbb{R}^m)$ for each m . Moreover, (A.6) implies $\bar{p}_{n,\theta}^{(r)}(s) \leq \bar{F}(s)$ for λ -a.e. s : conditional on (G_n, X_n) and on the focal-node-centered graph type $g(u)$, the law of $\bar{S}_{n,k}^{\theta,(r)}(u)$ is the pushforward of $L_u^{\theta,(r)}$ under $\pi_{Y,g(u)}$, and the corresponding λ -density is an $\text{Aut}_o(g(u))$ -average of the Lebesgue density $f_{B|\theta,u}^{(r)}$. Taking limits gives $\bar{p}_{\infty,\theta}^{(r)}(s) \leq \bar{F}(s)$ for λ -a.e. s . Therefore $|\bar{p}_{n,\theta}^{(r)}(s) - \bar{p}_{\infty,\theta}^{(r)}(s)| \leq 2\bar{F}(s)$ and dominated convergence yields $\|\bar{p}_{n,\theta}^{(r)} - \bar{p}_{\infty,\theta}^{(r)}\|_{L^1(\lambda)} \rightarrow 0$. Since $\bar{P}_{n,k}^{\theta,(r)}$ and $\bar{P}_{\infty,k}^{\theta,(r)}$ admit λ -densities, we have $\|\bar{P}_{n,k}^{\theta,(r)} - \bar{P}_{\infty,k}^{\theta,(r)}\|_{\text{TV}} = \frac{1}{2} \|\bar{p}_{n,\theta}^{(r)} - \bar{p}_{\infty,\theta}^{(r)}\|_{L^1(\lambda)}$ (Axler 2020, Theorem 9.10), and

hence

$$\|\bar{P}_{n,k}^{\theta,(r)} - \bar{P}_{\infty,k}^{\theta,(r)}\|_{\text{TV}} \longrightarrow 0. \quad (\text{A.11})$$

By (A.7) and the definition of $\bar{P}_{n,k}^{\theta,(r)}$,

$$\|\bar{P}_{n,k}^{\theta} - \bar{P}_{n,k}^{\theta,(r)}\|_{\text{TV}} \leq \frac{1}{n} \sum_{u \in V_n} \|L_u^{\theta} - L_u^{\theta,(r)}\|_{\text{TV}} \leq C_{\text{TV}} \rho^{r+1}, \quad (\text{A.12})$$

uniformly in n . Combining (A.11) and (A.12) yields

$$\limsup_{n \rightarrow \infty} \|\bar{P}_{n,k}^{\theta} - \bar{P}_{\infty,k}^{\theta,(r)}\|_{\text{TV}} \leq C_{\text{TV}} \rho^{r+1}. \quad (\text{A.13})$$

In addition, for each n the triangle inequality and (A.12) give $\|\bar{P}_{n,k}^{\theta,(r)} - \bar{P}_{n,k}^{\theta,(r')}\|_{\text{TV}} \leq C_{\text{TV}}(\rho^{r+1} + \rho^{r'+1})$, and letting $n \rightarrow \infty$ using (A.11) yields that for any $r, r' \geq 0$,

$$\|\bar{P}_{\infty,k}^{\theta,(r)} - \bar{P}_{\infty,k}^{\theta,(r')}\|_{\text{TV}} \leq C_{\text{TV}}(\rho^{r+1} + \rho^{r'+1}).$$

Thus $(\bar{P}_{\infty,k}^{\theta,(r)})_{r \geq 0}$ is Cauchy in total variation and converges to a limit, which we denote by $\bar{P}_{\infty,k}^{\theta}$. Taking $r \rightarrow \infty$ in (A.13) and using the Cauchy convergence of $\bar{P}_{\infty,k}^{\theta,(r)}$ yields (A.9). Uniqueness follows because total variation limits are unique.

For the uniform statement (A.10), note that (A.6) implies a uniform Lipschitz bound in θ in total variation. Indeed, total variation cannot increase under measurable pushforward, so for any $\theta, \theta' \in \Theta$,

$$\begin{aligned} \|\bar{P}_{n,k}^{\theta} - \bar{P}_{n,k}^{\theta'}\|_{\text{TV}} &\leq \frac{1}{n} \sum_{u \in V_n} \|L_u^{\theta} - L_u^{\theta'}\|_{\text{TV}} \\ &= \frac{1}{n} \sum_{u \in V_n} \frac{1}{2} \|f_{B|\theta,u} - f_{B|\theta',u}\|_{L^1} \\ &\leq \frac{1}{2} \tilde{C} \left(\sum_{m=1}^{d_k} \|F_m\|_{L^1} \right) \|\theta - \theta'\|. \end{aligned}$$

uniformly in n , where the last inequality follows by applying the mean value theorem pointwise in y and using the derivative bound in (A.6). The same Lipschitz bound holds for the limit family $\theta \mapsto \bar{P}_{\infty,k}^{\theta}$ by taking limits in (A.9). Let $L_{\text{TV}} := \frac{1}{2} \tilde{C} \sum_{m=1}^{d_k} \|F_m\|_{L^1}$ and fix $\eta > 0$. Choose a finite η -net $\{\theta^j\}_{j=1}^J$ of Θ . For any $\theta \in \Theta$ pick j with $\|\theta - \theta^j\| \leq \eta$ and use the triangle inequality together with the Lipschitz bound to obtain

$$\|\bar{P}_{n,k}^{\theta} - \bar{P}_{\infty,k}^{\theta}\|_{\text{TV}} \leq 2L_{\text{TV}}\eta + \max_{1 \leq j \leq J} \|\bar{P}_{n,k}^{\theta^j} - \bar{P}_{\infty,k}^{\theta^j}\|_{\text{TV}}.$$

Taking $\sup_{\theta \in \Theta}$, letting $n \rightarrow \infty$ (finite maximum), and then sending $\eta \downarrow 0$ yields (A.10).

Jensen–Shannon continuity and limiting separation.

Lemma 13 (Continuity of Jensen–Shannon divergence under total variation). *If $\|P_n - P\|_{\text{TV}} \rightarrow 0$ and $\|Q_n - Q\|_{\text{TV}} \rightarrow 0$, then $\text{JS}(P_n \| Q_n) \rightarrow \text{JS}(P \| Q)$.*

Proof of Lemma 13. Let $M_n := \frac{1}{2}(P_n + Q_n)$ and $M := \frac{1}{2}(P + Q)$ and define the dominating probability measure

$$\bar{\mu} := \frac{1}{2} \left(M + \sum_{m \geq 1} 2^{-m} M_m \right).$$

Then P_n, Q_n, P, Q, M_n, M are all absolutely continuous with respect to $\bar{\mu}$. Write p_n, q_n, p, q for the corresponding $\bar{\mu}$ -densities. Since $\|P_n - P\|_{\text{TV}} = \frac{1}{2} \|p_n - p\|_{L^1(\bar{\mu})}$ and $\|Q_n - Q\|_{\text{TV}} = \frac{1}{2} \|q_n - q\|_{L^1(\bar{\mu})}$ (Axler 2020, Theorem 9.10), we have $p_n \rightarrow p$ and $q_n \rightarrow q$ in $L^1(\bar{\mu})$ and hence also in $\bar{\mu}$ -measure. Define

$$\phi(a, b) := \frac{a}{2} \log \frac{2a}{a+b} + \frac{b}{2} \log \frac{2b}{a+b}, \quad \phi(0, 0) := 0.$$

Then $\text{JS}(P_n \| Q_n) = \int \phi(p_n, q_n) d\bar{\mu}$ and $\text{JS}(P \| Q) = \int \phi(p, q) d\bar{\mu}$. The function ϕ is continuous on $[0, \infty)^2$ and satisfies $0 \leq \phi(a, b) \leq (a+b) \log 2$. Since $p_n \rightarrow p$ and $q_n \rightarrow q$ in $L^1(\bar{\mu})$, the families $\{p_n\}_n$ and $\{q_n\}_n$ are uniformly integrable, and hence so is $\{p_n + q_n\}_n$ (Bogachev 2007, Proposition 4.5.3). The bound $0 \leq \phi(p_n, q_n) \leq (p_n + q_n) \log 2$ therefore implies that $\{\phi(p_n, q_n)\}_n$ is uniformly integrable as well. Since $(p_n, q_n) \rightarrow (p, q)$ in $\bar{\mu}$ -measure and ϕ is continuous, we have $\phi(p_n, q_n) \rightarrow \phi(p, q)$ in $\bar{\mu}$ -measure. By the Lebesgue–Vitali theorem (Bogachev 2007, Theorem 4.5.4), this yields $\phi(p_n, q_n) \rightarrow \phi(p, q)$ in $L^1(\bar{\mu})$, and hence $\int \phi(p_n, q_n) d\bar{\mu} \rightarrow \int \phi(p, q) d\bar{\mu}$.

Lemma 14 (Limiting criterion and separation). *Define the limiting reduced concentrated minimum-distance criterion*

$$\bar{V}_{\infty, k}^*(\theta) := -\log 4 + 2 \text{JS}(\bar{P}_{\infty, k}^{\theta_0} \| \bar{P}_{\infty, k}^{\theta}).$$

Under (P3), θ_0 is the unique minimizer of $\bar{V}_{\infty, k}^$ on Θ . Moreover, for every $\epsilon > 0$,*

$$\delta_{\infty}(\epsilon) := \inf_{\theta \in \Theta: \|\theta - \theta_0\| \geq \epsilon} (\bar{V}_{\infty, k}^*(\theta) - \bar{V}_{\infty, k}^*(\theta_0)) > 0. \quad (\text{A.14})$$

Proof of Lemma 14. By Lemma 12 and Lemma 13, the map $\theta \mapsto \bar{V}_{\infty, k}^*(\theta)$ is continuous on compact Θ . Faithfulness of Jensen–Shannon divergence (Lemma 4) implies that $\bar{V}_{\infty, k}^*(\theta) = -\log 4$ if and only if $\bar{P}_{\infty, k}^{\theta} = \bar{P}_{\infty, k}^{\theta_0}$, and (P3) then implies uniqueness of the minimizer at θ_0 . Continuity and compactness yield (A.14) by the extreme value theorem.

B Discussion

This appendix places the results of Sections 2 and 5 in broader perspective. It clarifies the substantive content of the assumptions (Section B.1), delineates the model classes covered by the identification and consistency arguments (Section B.2), discusses five methodological properties of the minimum-distance local-observation estimator (Section B.3), interprets the simulation evidence (Section B.4), and describes extensions suggested by the theory (Section B.5).

B.1 Assumption scope

The assumptions required for the identification and consistency results divide into four groups: graph architecture conditions (A1) and (A2), structural regularity conditions (B1) through (B3), identification conditions (D1) through (D3), and the additional large sample and sieve conditions (C1) through (C7) used in the consistency theorem. Each imposes a distinct restriction whose practical content is worth making explicit.

Bounded degree. Assumption (A1) requires that the maximum degree of G_n is bounded by a finite constant Δ for all n . This condition is responsible for four distinct roles in the analysis: it implies that focal-node-centered distance- k ego-neighborhood types form a finite set \mathcal{G}_k , which is necessary to construct the ego object space $\mathcal{S}_{n,k}$ (section 2.1); it bounds the maximal ball volume as $\nu_n(k) = O(\Delta^k)$, which controls the number of vertices that influence a single ego object; it bounds the chromatic number of the dependence-neighborhood graph in lemma 9(a), which is the key input to the effective number of independent local observations $m_{\text{eff},n}(r_n)$ in (3.3); and it bounds the overlap count of ego objects across focal nodes, which enters the memorization and batch overlap arguments of Section 4.2. The condition is natural in settings in which each vertex can maintain only a bounded number of direct relationships. It is restrictive in settings with highly heterogeneous degree sequences in which a small fraction of vertices accumulates a growing number of connections as n increases. Extending the framework to graphs with growing maximum degree requires controlling the rate at which $\nu_n(2(k + r_n))$ grows relative to n , which is the binding constraint in the effective number of independent local observations bound (3.3). If $\max_{u \in V_n} \deg_n(u) = O(n^\alpha)$ for some $\alpha \in [0, 1)$ and the truncation sequence satisfies $r_n = o(\log n)$, then $\nu_n(2(k + r_n)) = o(n)$ may still hold, leaving $m_{\text{eff},n}(r_n) \rightarrow \infty$; however, establishing this requires graph-specific degree concentration arguments beyond the bounded degree case.

Locality of interaction. Assumption (A2) requires that the i th coordinate of the equilibrium map depends on outcomes, covariates, and shocks only through the radius one neighborhood $B_{n,1}(i)$. This captures models in which each agent's response depends directly on its immediate contacts but not on more distant vertices except

through their propagated influence on immediate contacts. The propagation is handled by the equilibrium fixed point rather than by extending the interaction radius. The assumption is stated for interaction radius one but the paper notes (section 2.2) that the results extend to any fixed interaction radius $r_0 \geq 1$ by replacing $B_{n,1}$ with B_{n,r_0} throughout; the truncation and effective number of independent local observations conditions of Section 3 must then be revisited with the modified truncation ball geometry, as discussed further in Section B.5. This extension covers models such as games on networks in which each player’s payoff depends on the aggregate action of all players within two hops, or network formation models in which the decision to form a link depends on common neighbors.

Contraction. Assumption (B1) imposes a uniform ℓ_∞ contraction condition on the equilibrium map $y \mapsto T_{\theta,n}(y, X_n, \varepsilon_n)$ with modulus $\rho \in (0, 1)$. This condition serves as the single quantitative engine of the entire analysis: it implies uniqueness of the equilibrium via the contraction mapping theorem (lemma 1), it controls the rate at which shock influence decays with graph distance (lemma 7), and it determines the truncation depth sequence r_n required for consistency (theorem 2). The amplification factor $(1 - \rho)^{-1}$ in the sensitivity bound lemma 6(b) quantifies how the contraction modulus enters the Lipschitz dependence of equilibrium outcomes on structural parameters. In the linear in means model, the condition reduces to $|\beta| < 1$, the stability condition established in Section 5.1 for the equilibrium of that model. For general nonlinear models, verifying (B1) requires constructing an explicit bound on the ℓ_∞ operator norm of the state Jacobian $A_{\theta,n}(y, X_n)$ uniformly in y and θ , which in turn requires conditions on the parameterization of $h_{\theta,n}$.

Assumption (B1) rules out models with multiple equilibria, since any two distinct equilibria would contradict the contractivity. This is a substantive restriction in settings such as coordination games with strategic complementarities in which $\|A_{\theta,n}(y, X_n)\|_\infty \geq 1$ for some (y, θ) , so that (B1) fails and the amplification factor $(1 - \rho)^{-1}$ in lemma 6(b) cannot be bounded. Extending the framework to noncontractive settings would require either a selection rule that pins down a unique equilibrium for each parameter value or a partial identification approach that maps the structural parameter to a set of permissible distributions of local observations.

Additive shocks and exogeneity. Assumption (B2) requires that the equilibrium map factors as $T_{\theta,n}(y, x, \varepsilon) = h_{\theta,n}(y, x) + \varepsilon$, separating deterministic peer interactions from additive idiosyncratic shocks. Additivity enables the cancellation of shocks in differences that yields the contraction bound of lemma 1, and it enables the truncation coupling of Section 3, where two equilibria generated by shock vectors that agree on a ball can be compared coordinatewise by taking differences. State dependent noise structures, such as multiplicative heteroskedasticity of the form $\varepsilon_i \cdot \sigma(y_i, X_{n,i})$, break the additive separation and would require a modified truncation argument in which the shock difference no longer cancels from the contraction bound.

Assumption (B3) requires that, conditional on (G_n, X_n) , the shocks are i.i.d. across vertices with a strictly positive differentiable density. The i.i.d. property is used in two places: in the identification argument, where the product shock density yields the change of variables formula for the equilibrium density (Lemma 1’); and in the consistency argument, where the disjointness of truncation balls combined with i.i.d. shocks yields conditional independence within blocks (lemma 9(b)). The cross-sectional independence of shocks is conditional on the graph and covariates, which accommodates many forms of observable network correlation. What the assumption rules out is a residual network correlation in unobservables that persists after conditioning on (G_n, X_n) , such as correlated private information or common shocks to ego-level subsets of vertices.

Identification conditions. Assumptions (D1) through (D3) deliver the diffeomorphism from shocks to equilibrium outcomes and the positional heterogeneity required for identification. Assumption (D1) is a smoothness and locality condition on the state Jacobian and its mixed derivative with respect to θ , which underpins both the change of variables formula (Lemma 1’) and the unrolled derivative computation of Section 4.2. Assumption (D2) requires that a positive fraction of focal nodes are positioned such that their local outcome distributions respond nontrivially in every parameter direction, which is the variation that the adaptive test function exploits. Assumption (D3) requires that the map $u \mapsto \omega_u$ from focal nodes to observable signatures is injective, so that the mixture deconvolution step of theorem 1 can isolate unit-specific laws from the mixture distribution of local observations.

A failure of (D3) occurs on vertex transitive graphs with constant covariates, where all observable signatures ω_u coincide. This case is discussed explicitly in Section 5.1: identification from the k local law, if it holds at all, requires an argument that does not separate mixture components by signature. Whether and how identification can be achieved on vertex transitive graphs with constant covariates is an open question; any strategy addressing this case would need to rely on variation in the outcome component of the ego object across focal nodes, rather than on the joint (signature, outcome) structure, since all observable signatures coincide.

B.2 Model scope

The framework applies to any structural model satisfying assumptions (A1), (A2), (B1) through (B3), and (D1) through (D3). The interaction map $h_{\theta,n}$ is not required to be linear or even to admit a closed form; it suffices that $h_{\theta,n}$ is continuously differentiable in (y, θ) with a Jacobian satisfying the uniform contraction bound (D1)(ii). This permits parameterizations in which $h_{\theta,n}$ is a graph neural network whose weights are the structural parameter θ , provided that the network architecture enforces locality (A2) and the weight constraint enforces contractivity (B1). Examples satisfying the framework’s assumptions include: nonlinear peer effect models of the form

$Y_i = f_\theta(\sum_{j \in N(i)} W_{ij} Y_j, X_{n,i}) + \varepsilon_i$, where f_θ is contractively parameterised in its first argument; games on networks with unique Nash equilibria characterized as fixed points of contractive best response maps; and spatial autoregressive models with shocks that are not Gaussian and satisfy (B3).

Two categories of model fall outside the scope of the identification and consistency arguments developed here, in the sense that the assumptions used in Theorem 1 or Theorem 2 are violated and the present proofs do not apply without additional structure.

The first category consists of models in which all focal-node-centered ego objects have the same distribution under the observed equilibrium. When $P_{n,k}^{\theta_0,u}$ does not vary with u , the mixture distribution of local observations $P_{n,k}^{\theta_0}$ equals the common component law, and any parameter θ that produces the same common component law will match $P_{n,k}^{\theta_0}$. Identification then requires that the map $\theta \mapsto P_{n,k}^{\theta,u}$ is injective on the common component, not merely that different θ values produce different mixtures. Vertex transitive regular graphs with constant covariates are the leading example of this failure, as documented in Section 5.1. More generally, any graph sequence in which the number of distinct observable signatures $|\{\omega_u : u \in V_n\}|$ is bounded in n will produce a mixture distribution of local observations with a bounded number of components, limiting the identifying variation available to the criterion.

The second category consists of graph sequences on which the truncation decay condition (C4) and the effective number of independent local observations growth condition $m_{\text{eff},n}(r_n) \rightarrow \infty$ cannot be satisfied simultaneously by any truncation sequence (r_n) . The binding constraint, made precise in (3.3), is that $\nu_n(2(k+r_n)) = o(n)$ must hold for some r_n that also satisfies the decay requirement $\rho^{r_n} \|\varepsilon_n\|_\infty \rightarrow 0$ in probability. For a bounded degree Δ graph and Gaussian shocks, Section 5.1 shows that any $c \in (0, 1/(2 \log \Delta))$ yields $r_n = c \log n$ satisfying both conditions simultaneously. The window $c \in (0, 1/(2 \log \Delta))$ is nonempty for every finite Δ , so under bounded degree and Gaussian shocks the truncation decay and effective sample growth requirements in Theorem 2 are compatible. The failure mode arises when the maximum degree Δ_n grows with n : if Δ_n grows fast enough that $\nu_n(2(k+r_n)) \neq o(n)$ for every truncation sequence satisfying the decay condition, the effective number of independent local observations bound (3.3) does not diverge and the uniform law of large numbers of lemma 10 does not apply. On such sequences, whether identification from a fixed radius ego object is achievable without the growing sample of ego neighborhoods remains an open question.

B.3 Methodological merits

Five properties of the estimator deserve separate comment because each addresses a practical or theoretical limitation of existing alternatives.

Information efficiency of the complete distribution of local observations.

By lemma 3, the concentrated minimum-distance criterion equals a constant plus twice the Jensen–Shannon divergence between the observed and simulated k distributions of local observations $P_{n,k}^{\theta_0}$ and $P_{n,k}^\theta$. The Jensen–Shannon divergence is a divergence on the full distributions; it vanishes if and only if the two measures coincide (lemma 4). A moment based estimator that matches q selected statistics of the distribution of local observations uses a strict projection of the distributional information in $P_{n,k}^{\theta_0}$ rather than the measure itself. If the selected moments fail to separate θ from θ_0 , the moment estimator is not identified even when the full distributions of local observations are distinct. The minimum-distance criterion avoids this deficiency because the concentrated adaptive test function at parameter θ is the likelihood ratio $D^*(s) = dP_{n,k}^{\theta_0}/d(P_{n,k}^{\theta_0} + P_{n,k}^\theta)(s)$ (lemma 3). The map $s \mapsto D^*(s)$ is a monotone transformation of the likelihood ratio $dP_{n,k}^{\theta_0}/dP_{n,k}^\theta$, so any threshold test on $D^*(s)$ is equivalent to a threshold test on the likelihood ratio itself; the criterion therefore exploits all distributional contrast between the two distributions of local observations simultaneously. Parameters that are indistinguishable by any finite collection of prespecified moment conditions may still be separated by the criterion if their distributions of local observations differ on sets of positive measure.

Differentiability through equilibrium iteration without closed form derivatives.

The structural-parameter gradient $\nabla_\theta L_G(\theta; \phi)$ depends on θ through the simulated equilibrium \tilde{Y}_n^θ . Assumption (D1) requires only that the interaction map $h_{\theta,n}$ is continuously differentiable in (y, θ) with a Jacobian satisfying the uniform contraction bound. No closed form expression for the equilibrium map or its parameter derivative is required. The derivative $\partial_\theta \tilde{Y}_n^\theta$ is computed by reverse mode automatic differentiation through the unrolled Picard iteration, which records the computation graph of the iteration and applies the chain rule through the composition, as described in Section 4.2. The equilibrium approximation error satisfies $\|y^{(T)} - \tilde{Y}_n^\theta\|_\infty \leq \rho^T \|\tilde{Y}_n^\theta\|_\infty$, decaying at rate ρ^T , so that for any prescribed tolerance $\tau > 0$ a finite stopping depth $T = O(\log(1/\tau)/|\log \rho|)$ suffices for the equilibrium. For the derivative, in the linear case (equation (5.6)) the unrolled derivative error also decays at rate $|\beta|^T$ via the Neumann series. In the general nonlinear case, if additionally $y \mapsto A_{\theta,n}(y, X_n)$ and $y \mapsto \partial_\theta h_{\theta,n}(y, X_n)$ are uniformly Lipschitz in $\|\cdot\|_\infty$, Section 4.2 establishes a rate of $O(T\rho^T)$ for the derivative error; without this extra regularity, convergence holds qualitatively as $T \rightarrow \infty$ but a geometric rate is not guaranteed. This property extends the framework to structural models whose equilibrium map has no tractable analytical expression but can be evaluated by Picard iteration and whose interaction map $h_{\theta,n}$ is itself a differentiable function of θ , such as models parameterised by neural network interaction functions.

Flexible adaptive test function architecture for structured ego objects.

The adaptive test function D_ϕ operates on the ego object space $\mathcal{S}_{n,k}$, whose elements

are focal-node-centered induced subgraphs with node-level covariate-outcome vectors. This input format accommodates variable graph topology and variable covariate and outcome types without requiring the researcher to prespecify a fixed dimensional sufficient statistic. Message-passing neural networks, as noted in Section 4.3, are well suited to this format: they aggregate information from the k -hop neighborhood of the focal node in a manner that is invariant with respect to permutations of non-focal-node vertex labels, consistent with the focal-node-preserving isomorphism structure of $\mathcal{S}_{n,k}$. Node-level covariates and outcomes of mixed type—continuous covariates, categorical labels, and binary network positions—can be encoded as input node features and processed jointly, without requiring the researcher to select specific scalar summaries for each covariate type. The approximation condition (C7) requires only that the test-function sieve grows rich enough to approximate the optimal adaptive test function $D^*(s) = dP_{n,k}^{\theta_0}/d(P_{n,k}^{\theta_0} + P_{n,k}^{\theta})(s)$ uniformly in θ ; the specific architectural choices that satisfy (C7) depend on the structural model and the complexity of the likelihood ratio it implies.

A convergence diagnostic derived from theory. By lemma 3(b) and equation (5.8), the concentrated population criterion at the true parameter satisfies $V_{n,k}^*(\theta_0) = -\log 4$, and the population optimal adaptive test function satisfies $D^*(s) = 1/2$ for all $s \in \mathcal{S}_{n,k}$. These values are consequences of the general formulation and hold for any structural model satisfying assumptions (A1), (A2), (B1), and (B2), independently of the specific parameter value θ_0 and of the realized (G_n, X_n) . They imply that, at convergence, the discriminator loss \mathcal{L}_D stabilizes near $2 \log 2$ and the structural-parameter loss L_G stabilizes near $\log 2$. A researcher who observes both the adaptive test function and structural-parameter losses converging simultaneously to these theoretically determined values has empirical evidence that the algorithm has reached a neighborhood of the population minimizer, without requiring knowledge of θ_0 or access to an external benchmark. This diagnostic is an independent stopping criterion for the algorithm: it does not rely on validation data, a held-out subsample, or a parametric specification test, but derives directly from the structure of the Jensen–Shannon divergence at the population optimum.

Computational scalability. The computational cost of one forward pass is $O(Tn\bar{d})$, where T is the Picard stopping depth, n is the number of vertices, and \bar{d} is the mean degree, as stated in Section 4.2. Under bounded degree (A1), $\bar{d} \leq \Delta$, so the cost is linear in n for fixed Δ and T . The reverse pass for gradient computation has the same asymptotic order. Because ego extraction for each focal node costs $O(\Delta^k)$ independently of all other focal nodes given (G_n, X_n) , the forward pass and ego extraction steps decompose across focal nodes and can be parallelised across processor cores without requiring any between focal node communication. This structure is directly compatible with parallel execution on graphics processing units (GPUs), where the batch of ego objects is processed simultaneously on separate threads. For the Picard

iteration, each step evaluates $h_{\theta,n}$ at n vertices using only radius one neighborhood information under (A2), which maps to a sparse message passing operation on the graph adjacency structure and can be implemented with standard graph neural network primitives. Independent draws of the simulation shock vector $\tilde{\varepsilon}_n$ at the same θ are mutually independent conditional on (G_n, X_n) , so replicated equilibrium simulations at fixed θ require no shared computation and can be executed as parallel GPU streams.

B.4 Simulation interpretation

The linear in means model is chosen as the empirical illustration because it satisfies, or admits explicit verification of, every assumption required by the framework. Assumptions (A1), (A2), (B1) through (B3), and (D1) through (D3) hold as stated and are verified directly in Section C. The global Lipschitz condition (C3) does not hold for this model—because $\|h_{\theta,n}(y, X_n) - h_{\theta',n}(y, X_n)\|_\infty$ grows with $\|y\|_\infty$, which is unbounded over \mathbb{R}^n —but the weaker effective Lipschitz condition (C3') holds with $L_n = O(\sqrt{\log n})$ on the event \mathcal{A}_n of Section C, whose conditional probability tends to one, and the remaining consistency assumptions (C1), (C2), (C4) through (C7), and the uniform separation condition are verified in Appendix C. Section 5 then focuses on implications rather than assumption bookkeeping. The model also admits closed form expressions for the equilibrium density (equation (C.1)), the reduced form (section 5.1), and the truncation and effective number of independent local observations bounds (equation (5.4)), making it possible to trace the identification mechanism directly to the heterogeneity of the walk count structure $(I_n - \beta W_n)^{-1}$ across ego neighborhoods. The structural parameter (β, γ) has a transparent interpretation as the peer influence coefficient and the exogenous covariate effect.

The simulation reported in tables 1 to 3 and figures 1 to 3 provides finite sample evidence on the numerical behavior of the algorithm. The parameter trajectories in figure 1 converge to a stable neighborhood of $(\beta_0, \gamma_0) = (0.4, 1.5)$ from an initialisation $\theta^{(0)} \neq \theta_0$, with absolute errors at convergence of 0.0096 and 0.0046 respectively (table 2). Consistent with the convergence diagnostic of Section 5.2, the loss trajectories in figure 2 converge simultaneously to the theoretically predicted values $\mathcal{L}_D \rightarrow 2 \log 2$ and $L_G \rightarrow \log 2$. At the final iterate, the discriminator score distributions in figure 3 overlap near 1/2, indicating that the trained adaptive test function cannot reliably distinguish observed from simulated ego objects at the estimated parameter, which is the empirical counterpart of the population condition $D^*(\cdot) = 1/2$.

The graph used in the simulation has 249,813 vertices and 688,630 undirected edges, with mean degree 5.513 (table 1). The Picard iteration requires 15 steps to converge to a stopping tolerance, confirming that the logarithmic convergence rate $T = O(\log(1/\tau)/|\log |\beta||)$ is fast in practice for $|\beta| = 0.4$. The algorithm processes a network of this size without any implementation specialized to GPUs, demonstrating that the linear cost in n under bounded degree (A1) makes graphs of this scale

tractable; GPU execution would reduce wall clock time further by parallelising the ego extraction and Picard steps.

These three facts jointly confirm that the adversarial estimator is numerically stable, that its theoretical convergence criterion is operational as a stopping rule, and that it scales to network sizes relevant for empirical applications. They constitute an empirical illustration of the estimation procedure and should not be interpreted as a distributional assessment of the estimator across independent replications.

B.5 Future directions

Several assumptions of the present framework are amenable to relaxation, and each relaxation opens a distinct theoretical problem. The assumption that most immediately limits applicability to richer empirical settings is the i.i.d. shock condition (B3). Models in which agents within a community share a common background shock, or in which the link formation process is driven by unobservables correlated with outcomes, require assumptions beyond (B3). Extending the framework to allow for within community shock correlation requires replacing the i.i.d. property with a mixing condition that still permits the blocking and truncation argument of Section 3 to produce asymptotically independent summands. For block correlated shocks in which the block structure is observable, a conditional i.i.d. property within blocks may suffice; for general sparse correlation structures, the truncation argument would need to absorb the covariance across focal nodes into the effective number of independent local observations calculation.

The additivity condition (B2) restricts the shock entry to the equilibrium map to the additive form ε_i . Relaxing this to nonadditive shocks of the form $\sigma(Y_{n,i}, X_{n,i}, \varepsilon_i)$ introduces state dependence in the error variance. Under such a specification the contraction argument of lemma 1 requires that σ be Lipschitz in y with a constant L_σ satisfying $\rho + L_\sigma < 1$, so that the combined map $y \mapsto h_{\theta,n}(y, X_n) + \sigma(y, X_{n,i}, \varepsilon_i)$ remains contractive. The local approximation with influence decay of Section 3 would require a modified construction in which the difference between two equilibria generated by shock vectors agreeing on a ball depends on $L_\sigma \|Y - Y'\|_\infty$ rather than cancelling exactly.

A more substantial extension is to drop the unique equilibrium requirement by relaxing (B1) to noncontractive settings. In models with strategic complementarities in which $\|A_{\theta,n}(y, X_n)\|_\infty \geq 1$ for some (y, θ) so that assumption (B1) fails and multiple equilibria may arise, the structural parameter indexes a set of equilibria rather than a unique outcome, and identification requires specifying either a selection mechanism or a partial identification framework. The minimum-distance criterion as formulated in Section 2 compares the distribution $P_{n,k}^{\theta_0}$ of observed ego objects to the distribution $P_{n,k}^\theta$ of simulated ego objects under a chosen selection rule at θ . If the selection rule is correctly specified, the analysis of lemma 3 and theorem 1 extends directly. Partial identification in the network setting, where the parameter θ_0 is consistent with a set of

possible distributions of local observations, would require characterising the identified set of θ values compatible with equality of k distributions of local observations under some selection rule, a problem that combines the methods of the present paper with sharp identification analysis.

Finally, the interaction radius in assumption (A2) is fixed at one hop. Extending to games on networks in which the strategic interaction propagates through multiple hops—for example, network games in which each player’s payoff depends on the aggregate action within a radius of two or more hops—requires replacing (A2) with a locality condition at radius r_0 and adjusting the truncation and blocking constructions accordingly. The present analysis handles this extension at the level of notation (replacing $B_{n,1}$ with B_{n,r_0}), but the consistency theorem would require revisiting the effective number of independent local observations condition with the larger ego radius to confirm that the window of admissible truncation sequences r_n remains nonempty.

C Simulation exercise: assumption verification

This appendix verifies that the assumptions maintained in this paper apply to the linear-in-means model used in the simulations from Section 5. The structural model is

$$Y_n = \beta W_n Y_n + \gamma X_n + \varepsilon_n, \quad \varepsilon_{n,i} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

with interaction map

$$h_{\theta,n}(y, X_n) := \beta W_n y + \gamma X_n.$$

Assumptions are verified as follows.

Assumptions (A1) and (A2). Assumption (A1) is imposed on G_n directly: $\max_{u \in V_n} \deg_n(u) \leq \Delta < \infty$ for all n . Assumption (A2) holds because $(W_n y)_i = \sum_{j \in N(i)} W_{n,ij} y_j$ depends on y only through coordinates in $N(i) = B_{n,1}(i) \setminus \{i\}$, and

$$h_{\theta,n,i}(y, X_n) = \beta (W_n y)_i + \gamma X_{n,i}$$

depends on (y, X_n) only through coordinates indexed by $B_{n,1}(i)$.

Assumptions (B1) and (B2). Since $\|W_n\|_\infty \leq 1$ by construction,

$$\|h_{\theta,n}(y, X_n) - h_{\theta,n}(y', X_n)\|_\infty = |\beta| \|W_n(y - y')\|_\infty \leq |\beta| \|y - y'\|_\infty,$$

so (B1) holds with modulus $\rho = |\beta| < 1$. The decomposition (B2) holds with $h_{\theta,n}$ as defined above and ε_n entering additively.

Assumption (B3). The shocks $\varepsilon_{n,i} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ are i.i.d. conditional on (G_n, X_n) , independent of (G_n, X_n) , and admit the density

$$f_\varepsilon(z) = (2\pi)^{-1/2} \exp(-z^2/2),$$

which is $C^1(\mathbb{R})$, strictly positive everywhere, and satisfies $\int |f'_\varepsilon(z)| dz < \infty$.

Assumptions (D1)(i), (D1)(ii), and (D1)(iii). The state Jacobian is $A_{\theta,n}(y, X_n) = \beta W_n$, which is constant in y , satisfies $\|A_{\theta,n}\|_\infty = |\beta| \leq \rho < 1$ uniformly over all $y \in \mathbb{R}^n$ and $\theta \in \Theta$, and inherits locality from W_n : $(\beta W_n)_{ij} = 0$ for $j \notin B_{n,1}(i)$. Since $A_{\theta,n}$ does not depend on y , the mixed derivative $\partial_\theta A_{\theta,n}$ satisfies $\|\partial_\beta A_{\theta,n}\|_\infty = \|W_n\|_\infty = 1 < \infty$ and $\partial_\gamma A_{\theta,n} = 0$, so (D1)(iii) holds with $M_A = 1$.

Equilibrium density and smoothness. Since

$$F_{\theta,n}(y) = (I_n - \beta W_n)y - \gamma X_n$$

is affine in y with $\det \partial_y F_{\theta,n} = \det(I_n - \beta W_n)$ independent of y , the change of variables formula (2.2) gives

$$f_{Y|\theta,n}(y) = (2\pi)^{-n/2} |\det(I_n - \beta W_n)| \exp\left(-\frac{1}{2}\|(I_n - \beta W_n)y - \gamma X_n\|^2\right). \quad (\text{C.1})$$

This density is C^∞ in (θ, y) for all $(\theta, y) \in \Theta \times \mathbb{R}^n$. Fix n and θ , and set $A_{\theta,n} := I_n - \beta W_n$. Since $\|\beta W_n\|_\infty \leq |\beta| < 1$, the matrix $A_{\theta,n}$ is invertible. Let $s_{\min}(A_{\theta,n}) > 0$ denote its smallest singular value, so that $\|A_{\theta,n}y\| \geq s_{\min}(A_{\theta,n})\|y\|$ for all $y \in \mathbb{R}^n$. Then

$$\|A_{\theta,n}y - \gamma X_n\|^2 \geq \frac{1}{2}s_{\min}(A_{\theta,n})^2\|y\|^2 - \|\gamma X_n\|^2,$$

and every partial derivative $\partial_\theta^\alpha f_{Y|\theta,n}(y)$ is a polynomial in (y, X_n) times the same Gaussian kernel in (C.1). Therefore, for each multi-index α there exist a polynomial $p_{\alpha,n,\theta}$ and a constant $c_{n,\theta} > 0$ such that

$$|\partial_\theta^\alpha f_{Y|\theta,n}(y)| \leq p_{\alpha,n,\theta}(\|y\|) \exp(-c_{n,\theta}\|y\|^2),$$

and the right-hand side is integrable on \mathbb{R}^n for fixed n . Thus the hypotheses of a differentiation-under-the-integral theorem are satisfied and differentiation under the integral sign in (2.3) is valid for each fixed n (Conrad 2025, Theorem 12.3). Consequently, the conclusions of Lemma 1" and corollary 1 hold for the linear Gaussian model.

Remark on (C3). Assumption (C3) requires a deterministic finite constant L_θ such that

$$\|h_{\theta,n}(y, X_n) - h_{\theta',n}(y, X_n)\|_\infty \leq L_\theta \|\theta - \theta'\| \quad \text{for all } y \in \mathbb{R}^n.$$

For the linear model,

$$\|h_{\theta,n}(y, X_n) - h_{\theta',n}(y, X_n)\|_\infty \leq |\beta - \beta'| \|y\|_\infty + |\gamma - \gamma'| \|X_n\|_\infty,$$

and $\|y\|_\infty$ is unbounded over \mathbb{R}^n , so (C3) does not hold with a deterministic global constant. The consistency theorem of Section 3 is stated under the weaker effective Lipschitz condition (C3'), which we verify next.

Recall that the probability space in Section 3 includes the auxiliary outside copies $(\varepsilon_n^{(u,\text{out})})_{u \in V_n}$ and $(\tilde{\varepsilon}_n^{(u,\text{out})})_{u \in V_n}$. Define

$$\begin{aligned} \mathcal{A}_n := & \{ \|\varepsilon_n\|_\infty \leq C_* \sqrt{\log n} \} \cap \{ \|\tilde{\varepsilon}_n\|_\infty \leq C_* \sqrt{\log n} \} \\ & \cap \left\{ \max_{u \in V_n} \|\varepsilon_n^{(u,\text{out})}\|_\infty \leq C_* \sqrt{\log n} \right\} \cap \left\{ \max_{u \in V_n} \|\tilde{\varepsilon}_n^{(u,\text{out})}\|_\infty \leq C_* \sqrt{\log n} \right\}. \end{aligned}$$

For $Z \sim \mathcal{N}(0, 1)$, the Gaussian tail bound gives $\mathbb{P}(|Z| > t) \leq 2e^{-t^2/2}$ (Vershynin 2025, Proposition 2.1.2). A union bound over the at most $2n + 2n^2 \leq 4n^2$ Gaussian coordinates appearing in the shock collections in Section 3 yields $\mathbb{P}(\mathcal{A}_n^c) \rightarrow 0$ for sufficiently large fixed C_* , and hence $\mathbb{P}(\mathcal{A}_n \mid G_n, X_n) \rightarrow 1$.

On \mathcal{A}_n , for every focal node u and depth r , the truncated shocks $\varepsilon^{(u,r)}$ and $\tilde{\varepsilon}^{(u,r)}$ from lemma 8 satisfy $\|\varepsilon^{(u,r)}\|_\infty \leq C_* \sqrt{\log n}$ and $\|\tilde{\varepsilon}^{(u,r)}\|_\infty \leq C_* \sqrt{\log n}$. For any such shock vector ε ,

$$Y_n^\theta(\varepsilon) = (I_n - \beta W_n)^{-1}(\gamma X_n + \varepsilon),$$

so for any $\theta = (\beta, \gamma)$ and $\theta' = (\beta', \gamma')$ with $|\beta|, |\beta'| \leq \rho_{\max} < 1$, the resolvent identity gives

$$\|(I_n - \beta W_n)^{-1} - (I_n - \beta' W_n)^{-1}\|_\infty \leq \frac{|\beta - \beta'|}{(1 - \rho_{\max})^2}.$$

It follows that, on \mathcal{A}_n , there exists a constant $C < \infty$ depending only on $(\rho_{\max}, \|X_n\|_\infty)$ such that for all $\theta, \theta' \in \Theta$ and all u ,

$$\begin{aligned} & |\ell_{\theta,\phi}^{(u,r_n)}(u) - \ell_{\theta',\phi}^{(u,r_n)}(u)| \\ & \leq \frac{L_D}{\eta} \|Y_n^\theta(\tilde{\varepsilon}^{(u,r_n)}) - Y_n^{\theta'}(\tilde{\varepsilon}^{(u,r_n)})\|_{\infty, B_{n,k}(u)} \\ & \leq \frac{L_D}{\eta} C(1 + \sqrt{\log n}) \|\theta - \theta'\|. \end{aligned}$$

uniformly over $\phi \in \Phi_n$. Therefore (C3') holds with $L_n := \frac{L_D}{\eta} C(1 + \sqrt{\log n}) =$

$O(\sqrt{\log n})$.⁴²

Assumptions (C1), (C2), (C4)–(C7), and uniform separation. The parameter space is taken as

$$\Theta = \{(\beta, \gamma) : |\beta| \leq \rho_{\max} < 1, |\gamma| \leq M_\gamma\}$$

for fixed constants ρ_{\max} and M_γ . This specification satisfies (C1) and fixes the parameter bounds used in the linear-Gaussian verification.⁴³

⁴²Since $\Theta \subset \mathbb{R}^p$ is compact, for fixed $\epsilon > 0$ one has $\log N(\epsilon/L_n, \Theta, \|\cdot\|) = O(\log L_n) = O(\log \log n)$, which is $o(m_{\text{eff},n}(r_n))$ whenever $m_{\text{eff},n}(r_n) \rightarrow \infty$. Hence the entropy condition (3.4) holds under the same growth conditions on q_{n,r_n} and the test-function class.

⁴³For generic θ_0 , one has $\theta_0 \in \text{int}(\Theta)$. Assumption (C2) holds with $M_h = |\gamma| \|X_n\|_\infty$ under bounded covariates; (C4) is treated in Section 5.1; (C5) holds by metric definition; (C6)(i) is architecture-dependent through L_D ; and (C7) is imposed as sieve approximation. Uniform separation is derived from Proposition 1 under the additional limit conditions proved in Section A.