Single-Network Finite-Sample Inference in Strategic Network Formation Models*

Wayne Yuan Gao† Ming Li‡

September 23, 2026

Abstract

We develop a finite-sample valid inference procedure for strategic network formation models with endogenous network statistics, using only a single observed network. We impose no restrictions on network density, equilibrium selection, or the dependence structure induced by strategic interaction. Using a *bounding-by-c* technique, we obtain realization-wise sandwich inequalities whose middle term depends only on exogenous covariates and i.i.d. pairwise shocks. These inequalities deliver pathwise identifying restrictions and simulable finite-sample critical values in both parametric and semiparametric settings. Our proposed procedure is also computationally tractable: it does *not* involve solving, simulating, or enumerating equilibrium (sub)networks, and easily scales to networks with 10,000 agents in simulations. In two empirical applications, we find statistical evidence for positive link interdependence at the 95% confidence level.

Keywords: network formation, strategic interaction, single network, finite-sample inference, pairwise stability, partial identification

*We thank Giovanni Compiani, Áureo de Paula, Jiaying Gu, Lixiong Li, and seminar participants for helpful comments. Claude, ChatGPT and Refine.ink were used for research assistance.

†Department of Economics, University of Pennsylvania. Email: waynegao@upenn.edu

‡Department of Economics and Risk Management Institute, National University of Singapore. Email: mli@nus.edu.sg

1 Introduction

The formation of social and economic networks, such as friendships, information sharing, trade relationships, R&D alliances, and interbank lending, is shaped not only by the characteristics of the agents involved but also by the agents' strategic considerations over other network links. A link between two agents might be more attractive when they share common friends or when it connects them to well-positioned partners, for various reasons related to information, enforcement, and coordination. Analyzing network formation while explicitly accounting for strategic interdependence is a key problem in the econometrics of network formation (de Paula, 2020).

The same strategic interdependence that shapes network formation also poses several challenges for econometric inference, particularly when only a single network is observed. First, under pairwise stability (Jackson and Wolinsky, 1996), the mapping from model primitives to outcome networks is generally set-valued. Second, endogenous network statistics can induce complex network dependence. Third, conventional procedures are often computationally expensive: they often require equilibrium simulation or combinatorial enumeration of subnetwork configurations (e.g., Miyauchi, 2016; Sheng, 2020). Equilibrium multiplicity, complex dependence, and computational intractability are different manifestations of the same endogeneity problem induced by strategic interdependence. In particular, the realized network need not have a tractable reduced form, and statistics formed from its dyads need not satisfy familiar laws of large numbers or central limit theorems (cf. Menzel, 2021). Imposing a specific equilibrium selection rule or weak-dependence approximation could restore tractability, but would make inference depend on assumptions that are difficult to verify from a single realized network.

This paper, to the best of our knowledge, is the first in the econometrics literature to provide a *finite-sample valid inference* method on the structural parameters in a *strategic network formation* model under a *single-network* setting, together with the identification analysis that underlies it. The model we consider is the following:

$$Y_{ij} = \mathbf{1}\left\{Z_{ij}'\beta_0 + X_{ij}'\gamma_0 \geq \varepsilon_{ij}\right\},$$

where Z_{ij} is exogenous, ε_{ij} is an idiosyncratic pairwise shock, and $X_{ij} = \phi_{ij}(Y, Z)$ contains observed but endogenous network statistics, such as the number of common friends. Since X_{ij} involves realized links in the network, it can depend on every

pairwise shock in equilibrium. The parameter is $\theta_0 = (\beta_0', \gamma_0')'$, in which β_0 governs observed homophily and γ_0 governs strategic link interdependence.

The key device that enables us to handle the endogenous X_{ij} is a technique we call *bounding by c*. A formed link $Y_{ij} = 1$ reveals $\varepsilon_{ij} \leq Z_{ij}'\beta_0 + X_{ij}'\gamma_0$. If in addition the event that the index is at most a deterministic scanning threshold c, i.e., $Z_{ij}'\beta_0 + X_{ij}'\gamma_0 \leq c$, occurs, transitivity of the inequalities implies $\varepsilon_{ij} \leq c$, an event involving only the exogenous shock and the number c. Similarly, an absent link $Y_{ij} = 0$ reveals $\varepsilon_{ij} > Z_{ij}'\beta_0 + X_{ij}'\gamma_0$, so if in addition the index is at least c, we have $\varepsilon_{ij} > c$. Formed links thus generate lower-bound certificates for the frequency of the event $\{\varepsilon_{ij} \leq c\}$, while absent links generate the complementary upper-bound certificates. Averaging these event implications within cells of exogenous characteristics sandwiches each cell's empirical error distribution between two observable bounds. Importantly, the middle term of this sandwich depends only on Z and the i.i.d. pairwise errors, even though both bounds may depend arbitrarily on the equilibrium network through X. Aggregating over the threshold c converts these elementary bounds into a family of restrictions across the support of the latent index. The inequalities hold at every sample size, for every realization, and at every equilibrium.

A key contribution of this paper is to provide a *valid finite-sample inference* procedure using these realization-wise bounding-by-c inequalities. At the true parameter, our observable criterion is bounded by a control statistic involving only the exogenous covariates and pairwise shocks, and the distribution of this control statistic is exactly simulable, based on which we can construct semiparametric and parametric confidence sets with finite-sample coverage guarantees. In the parametric setup, the error distribution is specified and the control statistic can be simulated directly. We also develop sharper versions using constrained thresholding and Berk–Jones weighting (Berk and Jones, 1979). In the semiparametric setup, we show that it suffices to simulate uniform draws based on the probability integral transform, even though the error distribution is unknown. Under our procedure, testing a candidate parameter only requires dyad-level counting: there is no equilibrium solving, completion search, or graph-isomorphism matching. Furthermore, the finite-sample coverage result imposes no restriction on network density, equilibrium selection, or the dependence induced by strategic interaction. The cost of this robustness is conservatism: the control statistic bounds rather than reproduces the sampling distribution of the observable criterion.

The same inequalities also yield identifying restrictions under large-network

asymptotics, which clarifies the source of identification underlying our inference approach. Without sparsity or weak dependence assumptions, frequencies involving X_{ij} need not converge as networks grow large. We therefore formulate the identifying restrictions through *pathwise limit frequencies*: the limit superior of the observable lower bound and the limit inferior of the upper bound, both of which are allowed to remain random in the large-network limit, reflecting persistent indeterminacy from global dependence or equilibrium selection. Only the "middle sandwiched term", which consists solely of the exogenous covariates and i.i.d. shocks, converges to a deterministic limit and becomes the anchoring point for the identifying content. This differs from the usual population formulation of partial identification, which begins with deterministic (conditional) probabilities or moments. Here the observable envelopes may fluctuate indefinitely along the realized sequence, and (partial) identification is characterized by restrictions that hold between their pathwise limits. The result provides the population interpretation for the same inequalities used in our finite-sample inference procedure.

Our proposed procedures scale to networks of size $n = 10{,}000$ in our simulations. In the baseline simulation design with a scalar exogenous covariate, the parametric confidence set covers the truth in every replication (valid and conservative) and certifies the sign of the strategic coefficient (i.e., reject the null of $\gamma_0 = 0$ at 95% confidence level) from $n = 400$ onward. The fully semiparametric confidence set is also informative and certifies the sign in 97.8% of replications at $n = 3{,}200$, rising to 99.8% at $n = 6{,}400$ and $n = 10{,}000$. Additional simulations in the Supplemental Appendix show that confidence sets for γ_0 maintain coverage with additive common shocks and multivariate exogenous covariates, and that imposing symmetry on the error distribution improves sign certification.

We also apply our inference procedure to two real-world data sets, a high-school contact network ($n = 327$) and six Twitch friendship networks (up to $n = 9{,}498$). Across all networks in both applications, the projected 95% confidence sets for the strategic-interdependence coefficient lie strictly above zero, rejecting specifications without strategic interdependence. These findings highlight the importance of accounting for strategic interactions in models of network formation.

Related Literature

The papers most closely related to ours in the econometric literature are de Paula et al. (2018) and Menzel (2026). de Paula et al. (2018) obtain partial identification under bounded interaction distance and degree by classifying agents into local network types, defined by their local network structures along with node covariates. Their identification requires covariate and graph-isomorphism matching on local types, while their computation uses quadratic programs in a continuum-of-agents formulation. We instead use dyad-level threshold crossing, and impose no bound on degree, interaction depth, or density. Menzel (2026) also uses dyad-level frequencies, but restricts the endogenous covariates to a "node-incident" form that excludes fundamentally dyadic statistics such as common friends and derives a large-network Poisson approximation under additional error tail and equilibrium selection conditions. Our restrictions allow such dyadic statistics and are exact at the observed finite n.

The bounding-by-c technique itself originates outside the network setting: Gao and Wang (2026) introduce it for semiparametric identification of single-agent dynamic panel binary-choice models under partial stationarity.[1] They focus on a panel data setting with cross-sectional independence, while here we consider a very different network formation model where the endogenous covariate can be globally dependent. In a companion paper, Gao et al. (2026) combine the bounding-by-c approach with subnetwork differencing to accommodate individual fixed effects and focus on the identification of structural parameters. There the differencing annihilates node-level covariates entering as $Z_i + Z_j$, and the identification analysis there relies on population tetrad probabilities, whose existence in a single large network requires additional conditions on network dependence and equilibrium selection. We instead work without fixed effects, allow node-level covariates, and deliver valid finite-sample inference that invokes no population objects.

For the setting with many independent networks, Sheng (2020) studies similar strategic network formation models using bounds on subnetwork configurations and resolves multiplicity through the 0-1 bounds (Tamer, 2003). Her identifying restric-

[1] The underlying event-containment inequality is also related to the generalized instrumental-variable analysis of discrete outcomes (Chesher, 2010; Chesher and Rosen, 2017). Their bounds utilize the solution map and thus are able to obtain generally sharp identification. The bounding-by-c technique here instead formulates the bounds on the realized endogenous covariates directly without knowledge (or use) of the solution map, which affords tractability in our current strategic network formation setting, though we do not claim sharpness.

tions also require matching selected subnetwork configurations, and their asymptotic validity is established as the number of independent networks grows large. Gualdani (2021) instead makes a directed-link game tractable by decomposing it into local games. Both approaches complement our analysis of a single, potentially large network. Other approaches exploit monotone externalities (Miyauchi, 2016), model equilibrium selection (Mele, 2017; Badev, 2021), or use incomplete information to obtain conditional independence and two-step asymptotics (Leung, 2015; Wu, 2024; Ridder and Sheng, 2025; Comola and Dekel, 2026). For simultaneous discrete games with incomplete information, de Paula and Tang (2012) develop inference on the signs of interaction effects that is robust to equilibrium multiplicity; their many-market sampling environment differs from our single pairwise-stable network, but the sign of the strategic interaction effect is likewise the object our sign-revealing confidence sets target. Our approach delivers finite-sample inference from a single network without restricting equilibrium selection or the dependence induced by strategic interaction.

A separate line of literature considers dyadic network formation models with unobserved agent heterogeneity but no link interdependence (Graham, 2017; Jochmans, 2018; Dzemski, 2019; Gao, 2020; Auerbach, 2022a; Gao et al., 2023; Zeleneev, 2026; Li et al., 2026), where conditional independence of links across dyads can be exploited to deliver identification and asymptotic inference. Our problem is different: even conditional on observed and unobserved agent attributes, X_{ij} is a function of the realized network and can transmit equilibrium dependence across all dyads. The bounding-by-c technique is the leverage that allows us to retain the strategic channel.

Our inference results also contrast with single-network asymptotic theory based on weak dependence, exchangeability, or density restrictions (Leung, 2019; Kojevnikov et al., 2021; Menzel, 2021; Leung, 2022, 2023). These papers obtain laws of large numbers or central limit theorems by controlling how strongly distant observations can interact, by imposing an exchangeable-array representation, and/or by restricting the density regime. For strategic network formation, Leung and Moon (2026) derive a central limit theorem under subcritical spillovers and decentralized equilibrium selection. In contrast to this line of literature on network asymptotics, our inference procedures are derived directly from finite-sample properties of a control statistic, thus remaining valid when the global dependence generated by strategic interaction is present and when local network statistics do not converge to deterministic limits.

The closest econometric work on finite-sample inference is Rosen and Ura (2025).

Their main analysis considers the single-agent semiparametric binary-choice model under a conditional median restriction, which underlies the maximum score estimator of Manski (1975, 1985), and their inference likewise exploits a realization-wise inequality and simulable critical values. Here we consider multi-agent strategic network formation, where the endogenous network statistics are jointly determined with the entire equilibrium network and may therefore depend on all pairwise shocks, and use our bounding-by-c technique to handle the multi-agent strategic interdependence. Another closely related paper is Li and Henry (2026), who develop finite-sample inference in incomplete models with multi-agent interactions. Their procedure compares an optimal-transport statistic with a simulated worst-case statistic, obtained by optimizing over the equilibrium correspondence under a fully parametric specification. In contrast, our bounding-by-c technique yields closed-form dyad-level inequalities without simulating, solving, or optimizing over the equilibrium correspondence, which is prohibitively intractable in our strategic network formation context. In addition, our inference accommodates both parametric and semiparametric settings. Liu et al. (2026) use conformal inference under exchangeability to obtain prediction sets with finite-sample coverage for interval-valued outcomes. Their target is prediction rather than inference on structural parameters. Canay et al. (2017) similarly exploit group invariance in the form of approximate symmetry in an asymptotic framework, to construct randomization tests for structural parameters.

Some other inference procedures in network settings concern randomized or sharp-null causal hypotheses, or tests of hypotheses about the network distribution itself, rather than structural parameters under an unknown equilibrium (Athey, Eckles, and Imbens, 2018; Graham and Pelican, 2020; Pelican and Graham, 2022; Auerbach, 2022b). Such procedures exploit a known null randomization, asymptotic approximations to network test statistics, or a model that simplifies under the no-interaction null. Our confidence sets, by contrast, cover general values of the strategic interaction parameter and thus yield, by inversion, a test of no strategic interdependence. This test is a special case of the general specification test discussed in Remark 3.

The rest of the paper is organized as follows. Section 2 develops the model and identification analysis. Section 3 presents the finite-sample inference procedure. Sections 4 and 5 report simulations and applications. We conclude with Section 6. Proofs are available in the Appendix. The Supplemental Appendix provides additional implementation details along with supplemental simulation and theoretical results.

2 Model and Identification

2.1 Model Setup and Assumptions

Consider a set of n agents indexed by $i = 1, \ldots, n$, and an unweighted network among them represented by an $n \times n$ adjacency matrix $Y := (Y_{ij})_{i,j=1}^n$, where $Y_{ij} = 1$ if a link exists between agents i and j, and $Y_{ij} = 0$ otherwise. We focus on undirected networks, i.e., $Y_{ij} = Y_{ji}$ for all i, j, with $Y_{ii} = 0$. We index distinct unordered pairs as ij with $i < j$, and refer to each such pair ij as a *dyad*. Throughout the paper every sum $\sum_{ij}$ over dyads is understood to be $\sum_{i<j}$, so that each dyad is counted exactly once.

We consider the following network formation model with link interdependence

$$Y_{ij} = \mathbf{1}\left\{Z_{ij}'\beta_0 + X_{ij}'\gamma_0 \geq \varepsilon_{ij}\right\} \tag{1}$$

where $Z_{ij} := w_n(Z_i, Z_j)$ is a vector of exogenous covariates constructed from individual-level exogenous characteristics Z_i and Z_j via a known symmetric function w_n,[2] X_{ij} is a vector of potentially endogenous covariates, which may involve other links Y_{hk} with $(h, k) \neq (i, j)$, and ε_{ij} is an idiosyncratic pairwise shock. The exogenous covariates Z_{ij} can include a constant, homophily effects $|Z_i - Z_j|$ or $\mathbf{1}\{Z_i = Z_j\}$, level effects $Z_i + Z_j$, and interactions thereof. Level terms such as $Z_i + Z_j$ allow the model to capture differences in linking propensities associated with observed individual characteristics.[3] The structural parameters of interest are $\theta_0 := (\beta_0', \gamma_0')' \in \Theta$, and throughout this paper we use the shorthand notation

$$\delta_{ij} := \delta_{ij}(\theta_0), \quad \delta_{ij}(\theta) := Z_{ij}'\beta + X_{ij}'\gamma, \tag{2}$$

so that the link formation equation reads $Y_{ij} = \mathbf{1}\{\varepsilon_{ij} \leq \delta_{ij}\}$.

A key feature of our model is the presence of the endogenous covariates X_{ij}, which may arise as functions of the realized network Y, i.e.,

$$X_{ij} = \phi_{ij}(Y, Z) \tag{3}$$

where ϕ_{ij} is a known function mapping the realized network Y and the collection of

[2] We allow w_n to depend on the network size n, for example through the rescaling of latent positions (cf. Leung, 2019), to accommodate sparse designs in large networks.

[3] Our model excludes unrestricted unobserved individual effects, which our companion paper Gao et al. (2026) accommodates through subnetwork differencing.

individual covariate vectors $Z := (Z_i)_{i=1}^n$ to a vector of dyadic covariates. We collect these vectors in $X := (X_{ij})_{1 \le i < j \le n}$. We suppose that ϕ_{ij} does not depend on the realized own-link status Y_{ij},[4] but allow ϕ_{ij} to be both locally and globally defined.

One canonical local network statistic is the number of common friends of i and j, $\mathrm{CF}_{ij} := \sum_{k \neq i,j} Y_{ik} Y_{jk}$, which captures transitivity: the surplus of link ij may increase with the number of shared neighbors. The normalized version, $\overline{\mathrm{CF}}_{ij} := \frac{\mathrm{CF}_{ij}}{n-2} \in [0,1]$, is exactly the friends-in-common statistic of Sheng (2020), and a nonnegative coefficient on it delivers the nonnegative externalities exploited by Miyauchi (2016). Other leading choices of local network statistics are accommodated without modification: for example, the Jaccard index $J_{ij} := \mathrm{CF}_{ij} / \sum_{k \neq i,j} \mathbf{1}\{Y_{ik} + Y_{jk} \ge 1\}$ (with $J_{ij} := 0$ when the denominator vanishes), which measures overlap *relative* to the union of the two neighborhoods; second-order friends $\overline{\mathrm{SF}}_{ij} := \frac{1}{n-2} \sum_{k \neq i,j} (Y_{ik} + Y_{jk}) \in [0,2]$; and interactions with exogenous types, such as $\overline{\mathrm{CF}}_{ij} \cdot \mathbf{1}\{Z_i = Z_j\}$, which let the strength of transitivity vary with observables. Notably, our model also accommodates globally defined X_{ij}, i.e., those that cannot be pinned down completely by the local network structure around ij. Leading examples include various globally defined centrality measures, such as eigenvector centrality and closeness centrality (in the differenced form as described in Footnote 4). Such globally defined X_{ij} are usually highly nonlinear and implicit aggregations of links from the whole network Y, for which there exist no closed-form representations.

Because X_{ij} may depend on the realized network Y, equation (1) determines links jointly. The resulting network can be interpreted as a pairwise stable equilibrium of a strategic network formation game with transferable utility (cf. Jackson and Wolinsky, 1996; Sheng, 2020). In general, the game may have multiple equilibria for a given realization of primitives (Z, ε), where $\varepsilon := (\varepsilon_{ij})_{1 \le i < j \le n}$ collects all the idiosyncratic shocks. We represent the realized network abstractly by

$$Y \in g(Z, \varepsilon; \theta_0) \tag{4}$$

where g is the equilibrium correspondence. The equilibrium correspondence g gener-

[4]The own-link exclusion endows model (1) with the economic interpretation of pairwise stability, which rests on a *counterfactual* comparison between linking and not linking: any statistic $\tilde{\phi}_{ij}$ of the full network, such as the sum of the eigenvector centralities of i and j, enters through the counterfactual difference $\phi_{ij}(Y_{-ij}, Z) := \tilde{\phi}_{ij}((Y_{-ij}, 1), Z) - \tilde{\phi}_{ij}((Y_{-ij}, 0), Z)$, where (Y_{-ij}, y_{ij}) denotes the network with own-link state y_{ij}. If X_{ij} instead depends on the realized Y_{ij}, all our inference results still apply, but (1) loses its literal pairwise-stability interpretation.

ally lacks a tractable characterization. Even simple specifications of ϕ_{ij} can generate complex dependence through feedback among link decisions. The discrete, combinatorial space of possible networks further complicates the characterization of g. When ϕ_{ij} depends on features of the entire network, link decisions may also depend directly on distant links. Our identification approach does not require characterizing, evaluating, or computing g. As shown below, we extract identifying information from monotonicity restrictions that hold for every equilibrium, regardless of how complex the equilibrium correspondence is or how an equilibrium is selected.

We maintain the following assumptions throughout.

Assumption 1 (Model Specification)**.** *The observed data consist of one realization of* (Y, Z)*. For some fixed* $\theta_0 \in \Theta$*, equation* (1) *holds almost surely for every* $1 \leq i < j \leq n$*.*

Assumption 2 (IID Pairwise Errors)**.** *The components of* ε *are independently and identically distributed, with common CDF* F_ε*.*

Assumption 3 (Exogeneity of Z)**.** *The random collections* Z *and* ε *are independent.*

Assumptions 1–3 are standard in the strategic network formation literature. Under the transferable-utility interpretation of (1), Assumption 1 requires the observed network to be pairwise stable.[5] Assumption 2 requires the link-level surplus shocks to be independent and identically distributed across dyads but leaves their common distribution F_ε unrestricted at this stage. Our identification strategy applies in both parametric and semiparametric settings. We impose a parametric specification for F_ε only where needed.[6] Assumption 3 is an exogeneity condition on the observables Z. Importantly, we do *not* assume that X and ε are independent: since X is a function

[5] Assumption 1 also fixes the data environment we consider: a *single observed network*. The alternative environment of many independent networks (villages, schools, classrooms, markets) is more tractable, since conditional probabilities of observable events given the agents' exogenous characteristics are then directly identified and consistently estimable across networks, with no restriction on within-network dependence. Our restrictions apply to that case as well (Remark 2), but the single-network environment is the one our procedures are designed for.

[6] As is standard in binary-outcome models, the pair $(\theta_0, F_\varepsilon)$ is restricted only up to location and scale: adding a constant to the index and to ε, or multiplying both by a common $b > 0$, leaves (1) unchanged. A parametric specification such as the standard normal or logistic fixes both. In the semiparametric case we normalize the intercept to zero and fix the scale of one nonzero coefficient as $|\beta_1| = 1$, implemented as the union of the two sign charts $\beta_1 = \pm 1$ since a positive rescaling cannot change a coefficient's sign.

of the equilibrium network, it is endogenous and dependent on ε. This assumption also formalizes the naming convention of calling Z the exogenous covariates while calling X the endogenous covariates.

Remark 1 (Additive Common Shocks). An additive shock shared by all dyads can be absorbed into the free intercept in the parametric model, or into the unrestricted error location in the semiparametric model. The structural slope coefficients are unaffected. In the parametric case, the intercept represents the combined effect of the structural intercept and the realized common shock, which cannot be separately identified from a single network. Provided the remaining dyadic shocks satisfy our maintained assumptions conditional on the common shock, the finite-sample coverage guarantees apply to these parameters. Supplemental Appendix S.3.1 illustrates the parametric case and its validity via simulations.

2.2 The Bounding-by-c Event Inequalities

This subsection develops the bounding-by-c inequalities at the level of individual *events*, holding for every realization of the primitives, every equilibrium consistent with them, and every selection among equilibria. They underlie both the identifying restrictions of Section 2.3 and the finite-sample inference theory of Section 3.

Under model (1), each link satisfies the threshold-crossing condition $Y_{ij} = \mathbf{1}\{\varepsilon_{ij} \leq \delta_{ij}\}$. The key difficulty is that δ_{ij} contains the endogenous covariate X_{ij}. Its equilibrium value depends on the entire shock vector ε through an intractable equilibrium map g and an unknown selection mechanism.

The core idea of the bounding-by-c technique is to bound comparisons involving an endogenous threshold using a deterministic threshold c. Fix any constant $c \in \mathbb{R}$ and consider the event

$$Y_{ij} = 1 \quad \text{and} \quad \delta_{ij} \leq c.$$

Given that $Y_{ij} = \mathbf{1}\{\varepsilon_{ij} \leq \delta_{ij}\}$, together with the event $\{\delta_{ij} \leq c\}$, we can deduce from simple transitivity of the two inequalities that $\varepsilon_{ij} \leq c$, an event that involves only the exogenous shock and the deterministic threshold c. Similarly, the flipped event ($Y_{ij} = 0$ and $\delta_{ij} \geq c$) implies $\varepsilon_{ij} > \delta_{ij} \geq c$ and thus $\varepsilon_{ij} > c$. Formally, for any $c \in \mathbb{R}$,

$$Y_{ij}\mathbf{1}\left\{\delta_{ij} \leq c\right\} \;\leq\; \mathbf{1}\left\{\varepsilon_{ij} \leq c\right\} \leq 1 - (1 - Y_{ij})\,\mathbf{1}\left\{\delta_{ij} \geq c\right\}. \tag{5}$$

These are our core bounding-by-c sandwich inequalities.

Three remarks on the bounding-by-c technique are in order. First, the inequalities in (5) are valid algebraically for every realization of the primitives (Z, ε), the equilibrium network Y, and the induced network statistics X. Nothing about the equilibrium correspondence, its multiplicity, or the selection mechanism is used beyond model (1) itself. This is why the resulting restrictions are robust to equilibrium multiplicity and arbitrary, unknown selection.

Second, nonnegative weighted sums and, more generally, coordinatewise nondecreasing transformations preserve the inequalities in (5). Averaging within covariate cells underpins the inference procedures in Section 3. Taking conditional expectations yields analogous restrictions wherever those expectations are identified (Remark 2). The potential complex endogeneity of X_{ij} affects neither the validity, nor the computability, of such aggregations: no conditional distribution of X_{ij} given Z is ever modeled, estimated, or simulated.

Third, the test constant c generates a one-dimensional family of restrictions. Each c contributes one lower and one upper bound on $\mathbf{1}\{\varepsilon_{ij} \leq c\}$, whose probability is exactly the error CDF $F_\varepsilon(c)$. Hence, sweeping c over $\mathbb{R}$ is effectively a way to trace out bounds for the entire error CDF. The identifying content for θ comes from the interplay between the c-family, the aggregation across dyads, and the conditioning on the dyad's exogenous characteristics (Z_i, Z_j).

2.3 Identification under Large-Network Asymptotics

This subsection translates the realization-wise inequalities (5) into identifying restrictions under large-network asymptotics, in which the number of agents n grows to infinity. It should be emphasized that *none* of the finite-sample coverage guarantees of Section 3 relies on the identification analysis here, but it clarifies where the identifying power behind the inference procedure comes from.

Formally, we place all primitives on a single probability space: an infinite sequence $(Z_i)_{i\geq 1}$ of individual covariates, i.i.d. by Assumption 4 below, an array $(\varepsilon_{ij})_{1\leq i<j<\infty}$ of i.i.d. shocks independent of the covariate sequence, and an auxiliary randomization variable ξ independent of the two collections jointly (allowing randomized equilibrium selection). We write $Z^{(n)} = (Z_i)_{i=1}^n$ and $\varepsilon^{(n)} = (\varepsilon_{ij})_{1\leq i<j\leq n}$ for their finite restrictions. For each n, let $Y^{(n)} = g_n(Z^{(n)}, \varepsilon^{(n)}; \theta_0, \xi)$ be the realized network formed from the first n agents by the (n-dependent) equilibrium-and-selection mapping g_n,

which completes the equilibrium correspondence g in (4) with an arbitrary equilibrium selection mechanism under abstract (global) randomness in ξ. We write $\mathbb{P}$ for the joint law of $\big((Z_i)_{i\geq 1}, (\varepsilon_{ij})_{1\leq i<j<\infty}, \xi\big)$. The resulting equilibrium networks $Y^{(n)}$ need not be nested across n.

The Finite-Sample Sandwich

We now formulate the finite-sample conditional averages of the bounding-by-c inequalities in Section 2.2. We first define a discrete random variable $Z_{D,ij}$ below, which allows us to cover discrete, continuous, and mixed Z_i in a unified manner.

Definition 1 (Discretized Dyad Types/Cells)**.** Let $(\mathcal{Z}_{D,1}, \ldots, \mathcal{Z}_{D,\kappa})$ be an arbitrary finite, *symmetric*[7] partition of $\text{Supp}\,(Z_i, Z_j) \equiv \text{Supp}\,(Z_i)^2$ such that $\mathbb{P}\left((Z_i, Z_j) \in \mathcal{Z}_{D,k}\right) > 0$ for each $k = 1, \ldots, \kappa$. We define the discretized dyad type variable $Z_{D,ij}$ as $Z_{D,ij} := z_{D,k}$ if $(Z_i, Z_j) \in \mathcal{Z}_{D,k}$, where $z_{D,k}$ are κ distinct arbitrary cell labels. We write $\mathcal{Z}_D := \{z_{D,1}, \ldots, z_{D,\kappa}\}$ for the support of $Z_{D,ij}$.

When Z_i has continuous or mixed support, the above device allows us to proceed with a discretized dyad type, with *no loss of validity*: the finite-sample sandwich (9) below and the coverage guarantees of Section 3 hold for any partition as in Definition 1.[8] When Z_i has finite support, one can always take the finest symmetric partition, whose cells are the diagonal points $\{(z, z)\}$ and the symmetrized pairs $\{(z, z'), (z', z)\}$ for $z \neq z'$ in $\text{Supp}\,(Z_i)$. That said, in computation, one may still want to aggregate support points in $\text{Supp}\,(Z_i, Z_j)$ into coarser cells even when Z_i is discrete.

We start by fixing a candidate parameter θ, a threshold c, and a generic dyad type z_D in the support of $Z_{D,ij}$. Define the realized cell frequencies

$$\hat{p}_L^{(n)}(z_D, c; \theta) := \frac{\sum_{ij} Y_{ij} \mathbf{1}\{\delta_{ij}(\theta) \leq c\} \mathbf{1}\{Z_{D,ij} = z_D\}}{\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\}}, \tag{6}$$

$$\hat{p}_U^{(n)}(z_D, c; \theta) := 1 - \frac{\sum_{ij} (1 - Y_{ij}) \mathbf{1}\{\delta_{ij}(\theta) \geq c\} \mathbf{1}\{Z_{D,ij} = z_D\}}{\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\}}, \tag{7}$$

[7] Symmetric in the sense that $(z, z') \in \mathcal{Z}_{D,k}$ if and only if $(z', z) \in \mathcal{Z}_{D,k}$.

[8] The discretized type $Z_{D,ij}$ enters only as the conditioning variable through which dyads are grouped into cells. In the meanwhile, the model itself is *never discretized*, and $\delta_{ij}(\theta)$ is always evaluated at the raw covariates. Discretization therefore costs informational coarsening, not validity: the restrictions hold cell by cell for any partition, so a coarser partition simply delivers fewer of them. Coarsening can even be desirable in finite samples, since larger cells make the exogenous middle term concentrate more tightly and hence yield smaller critical values. See Section 3 for more details.

together with the *exogenous middle term*

$$\hat{F}_n(c; z_D) := \frac{\sum_{ij} \mathbf{1}\{\varepsilon_{ij} \le c\}\mathbf{1}\{Z_{D,ij} = z_D\}}{\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\}}. \tag{8}$$

All three objects are defined on the event that $\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\} \ge 1$, which occurs almost surely for all large n since $\mathbb{P}(Z_{D,ij} = z_D) = \mathbb{P}((Z_i, Z_j) \in \mathcal{Z}_{D,k}) > 0$. On the complementary event of an empty cell, we adopt the convention $(\hat{p}_L^{(n)}, \hat{F}_n, \hat{p}_U^{(n)}) := (0, 0, 1)$, which preserves the sandwich inequality below and, almost surely, affects only finitely many terms of each sequence. We emphasize that $\delta_{ij}(\theta) = Z_{ij}'\beta + X_{ij}'\gamma$ is still defined with the raw exogenous covariates Z_{ij}, *not* recomputed based on the discretized $Z_{D,ij}$. The frequencies $\hat{p}_L^{(n)}$ and $\hat{p}_U^{(n)}$ are computable from observed data (Y, X, Z), while the middle term $\hat{F}_n$ involves the unobserved shocks.

Averaging (5) within a given cell z_D therefore yields the following *finite-sample oracle sandwich*: for every n,

$$\hat{p}_L^{(n)}(z_D, c; \theta_0) \;\le\; \hat{F}_n(c; z_D) \;\le\; \hat{p}_U^{(n)}(z_D, c; \theta_0) \quad \text{for all } c \in \mathbb{R} \tag{9}$$

at every z_D such that $\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\} \ge 1$. The sandwich inequality (9) is stronger than a population moment inequality: it holds for each realization *before* any expectation is taken, and under arbitrary dependence in the realized network. All equilibrium (endogeneity) complications are confined to the two observable outer terms $\hat{p}_L^{(n)}$ and $\hat{p}_U^{(n)}$, while the middle term $\hat{F}_n$ contains only the exogenous Z and ε.

A Uniform Strong LLN for the Exogenous Middle Term $\hat{F}_n$

What drives our large-network identification result is the convergence of the middle term $\hat{F}_n(c; z_D)$ as $n \to \infty$. This convergence is the only place where we need the agents to be randomly sampled from a common population, which we now impose.

Assumption 4 (Random Sampling)**.** *The individual-level exogenous characteristics Z_i are independently and identically distributed across i.*

Assumption 4 is again a standard assumption in the network formation literature: in our paper it is used only in this subsection. None of the finite-sample inference results in Section 3 invokes it, where we condition on a finite n and the realized Z.

Lemma 1 (Uniform LLN for the Middle Term)**.** *Under Assumptions 2–4,*

$$\max_{z_D \in \mathcal{Z}_D} \sup_{c \in \mathbb{R}} \left| \hat{F}_n(c; z_D) - F_\varepsilon(c) \right| \longrightarrow 0, \quad \mathbb{P}\text{-almost surely.}$$

The proof is elementary after observing that the numerator and denominator of $\hat{F}_n$ are U-statistics of the jointly exchangeable, dissociated array formed by i.i.d. Z_i and ε_{ij}, so the strong law of large numbers for exchangeable arrays applies. The equilibrium and the endogenous X_{ij} play no role at all in this convergence.

Pathwise Identified Set

We now formulate identifying restrictions in the large-network limit. Define, for each (z_D, c, θ), the (possibly random) *pathwise limit frequencies*

$$p_L^\infty(z_D, c; \theta) := \limsup_{n \to \infty} \hat{p}_L^{(n)}(z_D, c; \theta), \quad p_U^\infty(z_D, c; \theta) := \liminf_{n \to \infty} \hat{p}_U^{(n)}(z_D, c; \theta), \tag{10}$$

which always exist. Crucially, the inequality $\limsup_n \hat{p}_L^{(n)} \leq \liminf_n \hat{p}_U^{(n)}$ does *not* follow from the finite-sample sandwich (9) alone, but it does follow once the middle term $\hat{F}_n(c; z_D)$ is shown to *converge* per Lemma 1: $\mathbb{P}$-almost surely,

$$\begin{aligned} p_L^\infty(z_D, c; \theta_0) &= \limsup_n \hat{p}_L^{(n)}(z_D, c; \theta_0) \leq \limsup_n \hat{F}_n(c; z_D) = F_\varepsilon(c), \\ p_U^\infty(z_D, c; \theta_0) &= \liminf_n \hat{p}_U^{(n)}(z_D, c; \theta_0) \geq \liminf_n \hat{F}_n(c; z_D) = F_\varepsilon(c), \end{aligned}$$

which yields the following.

Theorem 1. *Suppose Assumptions 1–4 hold. Then, $\mathbb{P}$-almost surely,*

$$p_L^\infty(z_D, c; \theta_0) \leq F_\varepsilon(c) \leq p_U^\infty(z_D, c; \theta_0) \quad \textit{for all } c \in \mathbb{R} \textit{ and all dyad types } z_D.$$

Consequently, the following results hold almost surely:

(i) Semiparametric case:

$$\max_{z_D \in \mathcal{Z}_D} p_L^\infty(z_D, c; \theta_0) \leq \min_{z_D} p_U^\infty(z_D, c; \theta_0) \quad \textit{for all } c \in \mathbb{R};$$

(ii) Parametric case: if furthermore F_ε is assumed to belong to a parametric family $F_\varepsilon(\cdot\,; \theta_\varepsilon)$, then the stacked true parameter $\overline{\theta}_0 := (\theta_0', \theta_{\varepsilon,0}')'$ satisfies

$$\max_{z_D \in \mathcal{Z}_D} p_L^\infty(z_D, c; \theta_0) \leq F_\varepsilon(c; \theta_{\varepsilon,0}) \leq \min_{z_D} p_U^\infty(z_D, c; \theta_0) \quad \textit{for all } c \in \mathbb{R}.$$

Accordingly, define the semiparametric *pathwise identified set*

$$\Theta_I^\infty := \left\{\theta \in \Theta : \max_{z_D \in \mathcal{Z}_D} p_L^\infty(z_D, c; \theta) \leq \min_{z_D} p_U^\infty(z_D, c; \theta) \text{ for all } c \in \mathbb{R}\right\},$$

so that Theorem 1(i) states $\theta_0 \in \Theta_I^\infty$ almost surely. The parametric counterpart is defined analogously from part (ii). The nonstandard formulation of this pathwise identification result warrants some discussion.

First, we clarify that we use the phrase "identified set" *without* claiming sharpness, and there is a substantive reason to doubt that sharp characterizations are feasible in our context. Sharp identified sets in games are often characterized by, in effect, *solving the model*: for each candidate parameter and realization of the latent primitives, enumerate the full set of equilibrium outcomes and check whether some selection reproduces the distribution of the observables. Such sharp identification results have been formalized through random-set and optimal-transport methods by Beresteanu, Molchanov, and Molinari (2011) and Galichon and Henry (2011). Here in the specific context of strategic network formation, the requirement of solving or inverting the pairwise-stability correspondence over a combinatorial outcome space is intractable in both theoretical and practical senses, and this intractability is exactly the motivation behind this paper. The value of the bounding-by-c restrictions lies in that they extract equilibrium-robust identifying content *without solving the game*, and inspire the finite-sample inference procedure in Section 3.

Second, Theorem 1 defines a *valid* identified set, but does *not* characterize the identified set as a *deterministic* population quantity as in standard microeconometric settings. Without weak-dependence conditions, $\hat{p}_L^{(n)}$ and $\hat{p}_U^{(n)}$ need not converge to deterministic quantities as $n \to \infty$: global interdependence, or an equilibrium selection that oscillates along the sequence, can keep them fluctuating, in which case p_L^∞ and p_U^∞, as well as the identified set Θ_I^∞, can remain *random*, with such unmodeled randomness abstractly captured by ξ. This marks an important departure from the standard microeconometric identification analysis, and we regard it as the right one in a single large-network setting. Conventional identification analysis, point- or set-valued, rests upon a deterministic feature of the observable distribution of data: a population conditional probability or moment, or its probability limit, which the data can in principle recover and upon which the model imposes restrictions. In our context no such object needs to exist. What is deterministically pinned down in the limit (or "population") is instead the *middle* term: the empirical distribution

of the exogenous shocks $\hat{F}_n$ converges to F_ε along almost every path, and all identifying restrictions come from sandwiching this convergent middle term between two possibly random observable bounds. The situation is reminiscent of laws of large numbers without ergodicity in time-series settings: for time-series processes, limits of intertemporal averages are in general random variables in the form of conditional expectations given the invariant σ-field rather than constants, and inference can still proceed given the realized time-series path. In the same spirit, identification here is defined *pathwise*, along the realized network sequence, without taking any stand on whether population limits of observable frequencies exist, or whether there exists a well-defined limit network model which all finite-n models converge to. Figure 1 depicts this pathwise formulation at a fixed (z_D, c). Along the realized network sequence, global dependence or an oscillating equilibrium selection can keep the observable envelopes fluctuating indefinitely, so their pathwise limits may remain random. Only the exogenous middle term $\hat{F}_n$ is guaranteed to converge, by Lemma 1, and it is this convergent term that anchors the identifying restrictions of Theorem 1.

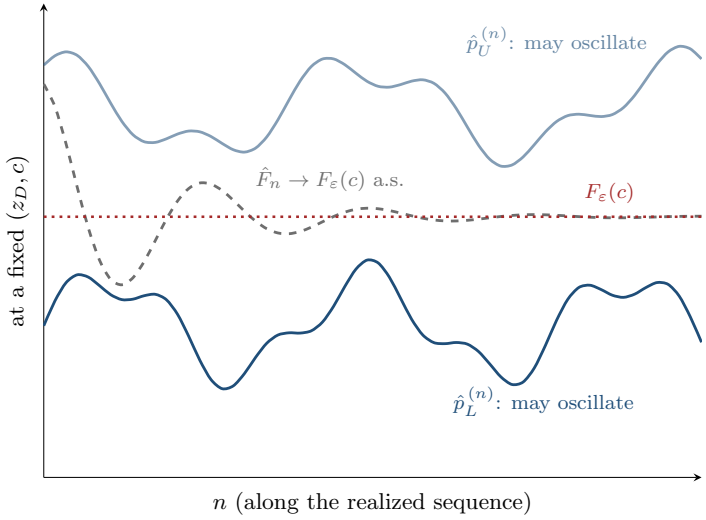

Figure 1: Illustration of Pathwise Limits

Remark 2 (The Many-Network Environment). When the researcher observes many independent networks drawn from a common data-generating process, everything above simplifies and the limit frequencies (10) are replaced by ordinary (conditional) population moments. Taking expectations of (5) conditional on the dyad type $Z_{D,ij} = z_D$, where the middle term equals $F_\varepsilon(c)$ free of Z by Assumptions 2 and 3, yields the

many-network analog of Theorem 1,

$$p_L(z_D, c; \theta_0) \;\leq\; F_\varepsilon(c) \;\leq\; p_U(z_D, c; \theta_0) \quad \text{for every } c \text{ and every dyad type } z_D,$$

with $p_L(z_D, c; \theta) := \mathbb{E}[Y_{ij}\mathbf{1}\{\delta_{ij}(\theta) \leq c\} \mid Z_{D,ij} = z_D]$, and $p_U(z_D, c; \theta) := 1 - \mathbb{E}[(1 - Y_{ij})\mathbf{1}\{\delta_{ij}(\theta) \geq c\} \mid Z_{D,ij} = z_D]$, which are directly identified and consistently estimable as the number of independent networks grows large, with *no* restriction on within-network dependence. Here, p_L and p_U are deterministic conditional probabilities at each given z_D, as in conventional microeconometric identification analysis.

3 Finite-Sample Inference in a Single Network

This section develops valid finite-sample inference procedures from the same realization-wise inequalities that underlie Theorem 1. The resulting confidence sets ("CSs"; singular "CS") have valid size control at the observed sample size, conditional on the realized Z with no large-network asymptotics involved.

3.1 Semiparametric Confidence Sets

The semiparametric identifying restriction in Theorem 1(i) can be encoded by

$$Q^\infty(\theta) := \sup_{c \in \mathbb{R}} \Big[\max_{z_D} p_L^\infty(z_D, c; \theta) - \min_{z_D} p_U^\infty(z_D, c; \theta) \Big]_+,$$

where $[x]_+ := \max\{x, 0\}$, so that $Q^\infty(\theta_0) = 0$ almost surely. Our procedure inverts a finite-sample analog of this criterion, and its logic consists of three steps. First, by the realization-wise sandwich (9), the observable criterion evaluated at the true θ_0 is bounded by a *control statistic* that involves only the exogenous (Z, ε), however complex the equilibrium network is. Second, by the probability integral transform, the conditional distribution of the control statistic (given Z) depends only on the cell sizes and not on the unknown F_ε, so it can be simulated exactly from uniform draws. Third, comparing the observable criterion against a simulated quantile of the control statistic yields a CS with valid finite-sample coverage. We develop each step below.

The Test Statistic and the Control Statistic

The finite-sample analog of Q^{∞} is

$$\widehat{Q}_n(\theta) := \sup_{c\in\mathbb{R}} \Big[\max_{z_D\in\hat{\mathcal{Z}}_D} \hat{p}_L^{(n)}(z_D, c; \theta) - \min_{z_D\in\hat{\mathcal{Z}}_D} \hat{p}_U^{(n)}(z_D, c; \theta) \Big]_+, \tag{11}$$

where $\hat{\mathcal{Z}}_D$ is the family of cells with at least $\underline{m}$ observations, i.e.,

$$\hat{\mathcal{Z}}_D := \big\{ z_D \in \mathcal{Z}_D : \hat{m}(z_D) \geq \underline{m} \big\}, \quad \hat{m}(z_D) := \sum_{i<j} \mathbf{1}\{Z_{D,ij} = z_D\}, \tag{12}$$

for a pre-chosen $\underline{m} \geq 1$.[9] Validity of our inference procedure is unaffected by any pre-chosen floor, again by the realization-wise validity of the bounding-by-c inequalities.[10] Also, though $\sup_{c\in\mathbb{R}}$ appears to be taken over the whole real line, it is actually computable with finitely many evaluations. This is because, at each fixed θ, the maps $c \mapsto \hat{p}_L^{(n)}(z_D, c; \theta)$ and $c \mapsto \hat{p}_U^{(n)}(z_D, c; \theta)$ are step functions whose jumps happen only at the finitely many realized index values in $\{\delta_{ij}(\theta) : i < j\}$, and under the $\leq$ / $\geq$ conventions in (6)–(7) the supremum of the violation is attained at one of these values. Hence, $\widehat{Q}_n$ is *exactly* computable at each candidate θ.

To control the finite-sample uncertainty in $\widehat{Q}_n(\theta_0)$, we use the "control statistic"

$$R_n := \sup_{c\in\mathbb{R}} \Big[\max_{z_D\in\hat{\mathcal{Z}}_D} \hat{F}_n(c; z_D) - \min_{z_D\in\hat{\mathcal{Z}}_D} \hat{F}_n(c; z_D) \Big], \tag{13}$$

which bounds $\widehat{Q}_n(\theta_0)$ by the following lemma.

Lemma 2 (Semiparametric Uncertainty Control)**.** *For every n, $\widehat{Q}_n(\theta_0) \leq R_n$.*

At the true parameter θ_0, the statistic $\widehat{Q}_n(\theta_0)$ on the left is observable and may carry arbitrary network endogeneity and dependence, while the control statistic R_n on the right contains no Y or X, and only involves the exogenous ε and Z. Any probability statement about R_n based on the independence assumption between ε and Z therefore translates directly into coverage guarantee for the truth θ_0.

[9] The choice of the floor $\underline{m}$ involves the following trade-off: raising $\underline{m}$ discards restrictions and thereby lowers the test statistic, but it also removes thin cells z_D that inflate the simulated critical value constructed below. In our experience the second effect typically dominates, so we impose a moderate floor rather than $\underline{m} = 1$.

[10] If no cell reaches the floor, so that $\hat{\mathcal{Z}}_D = \varnothing$, we define all test and control statistics (and their simulated copies) to be zero: the CSs below then equal the full candidate parameter space and the coverage statements hold trivially.

Simulability of the Control Statistic

It remains to obtain a critical value based on the control statistic R_n, whose definition still involves the unobserved shocks with unknown distribution F_ε, which we resolve using the probability integral transform. Conditional on Z, distinct cells contain disjoint sets of i.i.d. shocks (by Assumption 2). Let $U_{ij} := F_\varepsilon(\varepsilon_{ij})$ denote the transformed shocks and, for each cell, let G_{z_D} be their empirical CDF,

$$G_{z_D}(u) := \frac{1}{\hat{m}(z_D)} \sum_{i<j} \mathbf{1}\{U_{ij} \le u\}\mathbf{1}\{Z_{D,ij} = z_D\}, \quad u \in [0,1].$$

When F_ε is continuous, the middle term (8) satisfies, almost surely, $\hat{F}_n(c; z_D) = G_{z_D}\big(F_\varepsilon(c)\big)$ for all c: the transform maps each threshold c to the point $u = F_\varepsilon(c)$ on the uniform scale $[0,1]$. Substituting into (13), every value that the objective of R_n attains as c sweeps $\mathbb{R}$ is also attained by the same objective in u, so

$$R_n \;\le\; R^* := \sup_{u\in[0,1]} \Big[\max_{z_D \in \hat{\mathcal{Z}}_D} G_{z_D}(u) - \min_{z_D \in \hat{\mathcal{Z}}_D} G_{z_D}(u) \Big],$$

which holds with exact equality when F_ε is continuous, as shown in the proof of Theorem 2. The key observation is that the U_{ij} are i.i.d. Uniform$[0,1]$ by the probability integral transform (again for continuous F_ε), so, conditional on Z, the distribution of R^* depends on the primitives *only through the cell sizes* $\{\hat{m}(z_D)\}$: it is the distribution of the same range statistic computed from independent uniform samples of those sizes, whatever the unknown F_ε may be. When F_ε has atoms, the randomized distributional transform still yields i.i.d. uniform variables and preserves the almost-sure bound $R_n \le R^*$. The same uniform simulation therefore remains valid, although the bound may be strict and introduce additional conservatism.

We simulate the exact conditional distribution of R^* as follows. For $b = 1, \dots, B$, draw $\hat{m}(z_D)$ i.i.d. uniform random variables on $[0,1]$ for each $z_D \in \hat{\mathcal{Z}}_D$, independently across b and across cells, and independently of the observed data and of the realized shocks. Let $G^{(b)}_{z_D}$ be the empirical distribution function of the draws in cell z_D, the simulated counterpart of G_{z_D}, and set

$$R^{(b)} := \sup_{u\in[0,1]} \Big[\max_{z_D \in \hat{\mathcal{Z}}_D} G^{(b)}_{z_D}(u) - \min_{z_D \in \hat{\mathcal{Z}}_D} G^{(b)}_{z_D}(u) \Big].$$

Conditional on Z, $R^*, R^{(1)}, \dots, R^{(B)}$ are i.i.d. Define $\hat{q}^R_{n,1-\alpha}(Z)$ as the k-th smallest value among $R^{(1)}, \dots, R^{(B)}$, where $k := \lceil (1-\alpha)(B+1) \rceil$. Exchangeability ensures

conditional size control at level α for every finite B with $k \leq B$.[11]

The Semiparametric Confidence Set and its Validity

We then construct the semiparametric CS as

$$\widehat{\mathcal{C}}_n^{\text{semi}}(1-\alpha) := \big\{\theta : \ \widehat{Q}_n(\theta) \leq \hat{q}^R_{n,1-\alpha}(Z)\big\}. \tag{14}$$

Probability statements for CSs based on simulated critical values are understood to be taken over the Monte Carlo draws used to construct those critical values as well. Note that the inference procedure uses only the cell sizes $\hat{m}(z_D)$, *without* the need to draw ε from the unknown distribution F_ε or to simulate any network formation process. The power of the test comes from cross-cell contradictions: because every cell's shocks are drawn from the same F_ε, at θ_0 the cell envelopes must overlap up to oracle sampling noise, whereas an incompatible θ opens cross-cell gaps too large to be oracle noise, and the simulated critical value calibrates exactly this noise scale from the cell sizes $\hat{m}(z_D)$.

Theorem 2 (Semiparametric Confidence Set)**.** *Suppose Assumptions 1–3 hold. Then, for every n and every B such that $\lceil (1-\alpha)(B+1) \rceil \leq B$,*

$$\mathbb{P}\big(\theta_0 \in \widehat{\mathcal{C}}_n^{\text{semi}}(1-\alpha) \mid Z\big) \geq 1-\alpha.$$

The coverage guarantee also holds unconditionally. Theorem 2, like every other finite-sample result below, conditions on the realized Z and therefore does not invoke Assumption 4. The proof formalizes the domination and exchangeability arguments developed above. An analytic alternative to the simulated critical value, based on the Dvoretzky–Kiefer–Wolfowitz inequality, together with the rate at which the critical value shrinks, is provided in Supplemental Appendix S.4 (Proposition S.1).

3.2 Parametric Confidence Sets

We now consider the parametric case, in which the true error CDF is $F_\varepsilon(\cdot\,;\theta_{\varepsilon,0})$ for an unknown finite-dimensional parameter $\theta_{\varepsilon,0}$. We write $\overline{\theta} := (\theta', \theta_\varepsilon')'$ for a candidate stacked parameter, as in Theorem 1(ii). The logic of Subsection 3.1 carries over,

[11] This is the classical Monte Carlo test construction of Dwass (1957) and Barnard (1963), and is also discussed in Dufour (2006). The requirement $\lceil (1-\alpha)(B+1) \rceil \leq B$ is mild: at $\alpha = 0.05$ the smallest admissible value is $B = 19$. We use $B = 999$ in our simulations and empirical applications.

with one simplification that shapes the whole construction: the middle term of the sandwich has a known functional form. At each candidate $\overline{\theta}$, we compare the observable envelopes directly with $F_\varepsilon(\cdot\,;\theta_\varepsilon)$ and simulate the control statistic under the corresponding null by drawing shocks from that distribution.

Note that, when F_ε is taken to be normal or logistic (as commonly adopted in the literature), it is without loss of generality to set F_ε to be standard normal or standard logistic as location and scale normalization, so that no free parameter θ_ε is required.[12] That said, we keep θ_ε for theoretical generality.

Specifically, we define the finite-sample test statistic

$$\widehat{Q}_n^{\mathrm{par}}(\overline{\theta}) := \sup_{c\in\mathbb{R}}\Big[\max\Big\{\max_{z_D\in\hat{\mathcal{Z}}_D}\hat{p}_L^{(n)}(z_D,c;\theta)-F_\varepsilon(c;\theta_\varepsilon),\; F_\varepsilon(c;\theta_\varepsilon)-\min_{z_D\in\hat{\mathcal{Z}}_D}\hat{p}_U^{(n)}(z_D,c;\theta)\Big\}\Big]_+. \tag{15}$$

Correspondingly, we define the control statistic

$$R_n^{\mathrm{par}}(\theta_\varepsilon) := \max_{z_D\in\hat{\mathcal{Z}}_D}\sup_{c\in\mathbb{R}}\big|\hat{F}_n(c;z_D)-F_\varepsilon(c;\theta_\varepsilon)\big|, \tag{16}$$

which controls the uncertainty in $\widehat{Q}_n^{\mathrm{par}}(\overline{\theta})$ at the truth.

Lemma 3 (Parametric Uncertainty Control)**.** *Suppose $F_\varepsilon = F_\varepsilon(\cdot\,;\theta_{\varepsilon,0})$ is known up to θ_ε. Then, for every n, $\widehat{Q}_n^{\mathrm{par}}(\overline{\theta}_0) \le R_n^{\mathrm{par}}(\theta_{\varepsilon,0})$.*

For each candidate θ_ε, we simulate the corresponding null distribution of $R_n^{\mathrm{par}}(\theta_\varepsilon)$ to obtain the critical value. Specifically, for $b=1,\dots,B$, draw (independently across b, and independently of the observed data) an i.i.d. pair-shock array $\{\varepsilon_{ij}^{(b)}\}_{ij}$ from $F_\varepsilon(\cdot\,;\theta_\varepsilon)$, compute $\hat{F}_n^{(b)}(c;z_D)$ and then

$$R_n^{\mathrm{par},(b)}(\theta_\varepsilon) := \max_{z_D\in\hat{\mathcal{Z}}_D}\sup_{c\in\mathbb{R}}\big|\hat{F}_n^{(b)}(c;z_D)-F_\varepsilon(c;\theta_\varepsilon)\big|.$$

Let $\hat{q}_{n,1-\alpha}(Z;\theta_\varepsilon)$ be the $\lceil(1-\alpha)(B+1)\rceil$-th smallest value of the simulated statistics $R_n^{\mathrm{par},(b)}(\theta_\varepsilon)$ where $b=1,\dots,B$ similarly as before. We define the parametric CS as

$$\widehat{\mathcal{C}}_n^{\mathrm{par}}(1-\alpha) := \big\{\overline{\theta}:\ \widehat{Q}_n^{\mathrm{par}}(\overline{\theta}) \le \hat{q}_{n,1-\alpha}(Z;\theta_\varepsilon)\big\}.$$

Theorem 3 (Parametric Confidence Set)**.** *Suppose Assumptions 1–3 hold with $F_\varepsilon = F_\varepsilon(\cdot\,;\theta_{\varepsilon,0})$ known up to the finite-dimensional parameter $\theta_{\varepsilon,0}$. Then, for every n and*

[12] When F_ε is set to be standard normal or standard logistic, a constant term should be included in Z_{ij}, and all components of θ should be left unrestricted.

every B *such that* $\lceil (1-\alpha)(B+1) \rceil \leq B$,

$$\mathbb{P}\left(\overline{\theta}_0 \in \widehat{\mathcal{C}}_n^{\text{par}}(1-\alpha) \,\middle|\, Z\right) \geq 1-\alpha.$$

Theorem 3 parallels Theorem 2, with two differences worth noting. First, knowledge of F_ε strengthens the restrictions: the parametric criterion (15) tests every cell against F_ε two-sidedly, whereas the semiparametric criterion (11) can exploit only cross-cell contradictions between the lower and upper bounds. Correspondingly, the parametric CSs are tighter in the simulations in Section 4. Second, the critical value is simulated by drawing shocks from the specified $F_\varepsilon(\cdot\,;\theta_\varepsilon)$ rather than from the uniform distribution, and it therefore varies with the candidate θ_ε, though not with θ.

Remark 3 (Overidentification and Specification Testing). Set identification does not preclude specification testing. Let $\widehat{\mathcal{C}}_n$ denote a $1-\alpha$ confidence set from this section, constructed under the conditions of its coverage theorem, and fix a candidate parameter set Θ_{cand}. Consider the null that the model and its maintained assumptions, including any restriction on the error distribution, hold for some parameter in Θ_{cand}. Reject this null when $\widehat{\mathcal{C}}_n \cap \Theta_{\text{cand}} = \varnothing$. Under the null, the true parameter belongs to Θ_{cand}, so an empty intersection excludes it from the confidence set. The corresponding coverage guarantee therefore gives

$$\mathbb{P}\Big(\widehat{\mathcal{C}}_n \cap \Theta_{\text{cand}} = \varnothing \,\Big|\, Z\Big) \leq \alpha.$$

Taking Θ_{cand} to be the full parameter space tests the model itself; a smaller set tests the model jointly with the imposed parameter restrictions.

3.3 Improved Confidence Sets

The sandwich (9) is a *collection* of pointwise inequalities, one per cell z_D and threshold c, so any transformation of both sides that preserves the inequalities yields another valid test statistic. This freedom lets us account for the different amounts of finite-sample uncertainty carried by different (z_D, c) configurations, which leads to a less conservative test and hence a tighter CS. Here we focus on the *parametric case*, and leave the semiparametric counterpart to Supplemental Appendix S.1.[13]

[13]The refinement used in the semiparametric columns of our simulations studentizes each cross-cell gap $\hat{p}_L^{(n)}(z_D, c; \theta) - \hat{p}_U^{(n)}(z_D', c; \theta)$ by $e(z_D) + e(z_D')$, where $e(z_D) \propto \hat{m}(z_D)^{-1/2}$ is the cell's Dvoretzky–Kiefer–Wolfowitz radius. The resulting range is compared with its own $\lceil (1-\alpha)(B+1) \rceil$-th smallest simulated value from the uniform draws of Section 3.1, which discounts gaps involving thin cells and

At a candidate $\overline{\theta}$, define

$$V^{+}(z_D, c; \overline{\theta}) := \left[\hat{p}_L^{(n)}(z_D, c; \theta) - F_\varepsilon(c; \theta_\varepsilon)\right]_+, \quad R^{+}(z_D, c; \theta_\varepsilon) := \left[\hat{F}_n(c; z_D) - F_\varepsilon(c; \theta_\varepsilon)\right]_+,$$
$$V^{-}(z_D, c; \overline{\theta}) := \left[F_\varepsilon(c; \theta_\varepsilon) - \hat{p}_U^{(n)}(z_D, c; \theta)\right]_+, \quad R^{-}(z_D, c; \theta_\varepsilon) := \left[F_\varepsilon(c; \theta_\varepsilon) - \hat{F}_n(c; z_D)\right]_+.$$

At the truth $\overline{\theta}_0$, the sandwich (9) can be written as $V^{\pm}(z_D, c; \overline{\theta}_0) \leq R^{\pm}(z_D, c; \theta_{\varepsilon,0})$ for every $(z_D, c, \pm)$, which we abbreviate to $V\left(\overline{\theta}_0\right) \leq R\left(\theta_{\varepsilon,0}\right)$ with V, R understood as the concatenations across all such combinations. Hence, for any *nondecreasing*[14] *aggregation* function T mapping the space of V or R into a scalar, we have

$$T\left(V\left(\overline{\theta}_0\right)\right) \;\leq\; T\left(R\left(\theta_{\varepsilon,0}\right)\right). \tag{17}$$

Taking T to be the maximum over all $(z_D, c, \pm)$ combinations recovers $\widehat{Q}_n^{\text{par}}$ and R_n^{par} from Section 3.2, in which case (17) specializes to Lemma 3. More generally, fix any *known* measurable nondecreasing $T_{Z,\overline{\theta}}$, which may depend freely on $\left(Z, \overline{\theta}\right)$ but not on (Y, X), the realized shocks, or the Monte Carlo draws. Simulating B i.i.d. copies $R^{(1)}, \ldots, R^{(B)}$ from Z and $F_\varepsilon\left(\cdot\,; \theta_\varepsilon\right)$ exactly as in Section 3.2, and letting $\hat{c}_{n,1-\alpha}^{T}(Z; \overline{\theta})$ be the $\lceil (1-\alpha)(B+1) \rceil$-th smallest value of the aggregated draws $T_{Z,\overline{\theta}}(R^{(b)})$, the T-induced CS below is again finite-sample valid:

$$\widehat{\mathcal{C}}_n^{T}(1-\alpha) := \left\{\overline{\theta} :\; T_{Z,\overline{\theta}}\left(V\left(\overline{\theta}\right)\right) \leq \hat{c}_{n,1-\alpha}^{T}\left(Z; \overline{\theta}\right)\right\}.$$

Theorem 4 (Valid Parametric Confidence Sets under Nondecreasing Aggregation)**.** *Suppose Assumptions 1–3 hold with $F_\varepsilon = F_\varepsilon(\cdot\,; \theta_{\varepsilon,0})$ known up to the finite-dimensional parameter $\theta_{\varepsilon,0}$. Then, for every n and every B such that $\lceil (1-\alpha)(B+1) \rceil \leq B$,*

$$\mathbb{P}\left(\overline{\theta}_0 \in \widehat{\mathcal{C}}_n^{T}(1-\alpha) \,\middle|\, Z\right) \geq 1 - \alpha.$$

The choice of T matters because it governs how uncertainty is weighted across $(z_D, c, \pm)$, and hence the distribution of $T(R(\theta_{\varepsilon,0})) - T\big(V(\overline{\theta}_0)\big)$. The maximum used in Lemma 3 is one of the simplest and most intuitive encoding of the identifying restrictions in Theorem 1, but in finite samples it can suffer from high uncertainty: one thin cell at an uninformative threshold can become the maximum by chance and move the test statistic or the critical value drastically. A better aggregation limits the influence of any single $(z_D, c, \pm)$ configuration that only has very few data points.

prevents one thin cell's sampling noise from setting the allowance for every pair.

[14] That is, $T\left(v\right) \leq T\left(v'\right)$ whenever $v \leq v'$ elementwise.

Throughout Sections 4 and 5, we implement the following two aggregation improvements: *constrained thresholding* and *Berk–Jones weighting*. We provide a short description below to convey the key ideas underlying these improvements, while more details of the exact implementation are available in Supplemental Appendix S.1.

First, we work with a restricted set of test thresholds c, which we call *constrained thresholding*. At each (θ, Z) we keep only those c lying in the *a priori* range of the candidate index in cell z_D,

$$C(z_D;\theta,Z)=\Big[\inf_{\tilde{z}:\,\tilde{z}_D=z_D} w_n\,(\tilde{z})'\,\beta+\inf_{x\in\mathcal{X}}\gamma' x,\quad \sup_{\tilde{z}:\,\tilde{z}_D=z_D} w_n\,(\tilde{z})'\,\beta+\sup_{x\in\mathcal{X}}\gamma' x\Big],$$

where $\mathcal{X}$ is a known deterministic set containing every value X_{ij} can take, so that the realized X never enters and $C(z_D;\theta,Z)$ is (θ,Z)-measurable. Outside this window neither $\hat{p}_L^{(n)}$ nor $\hat{p}_U^{(n)}$ varies with c, so those coordinates say nothing about θ, while their oracle counterparts $R^{\pm}$ keep inflating the simulated maximum. Discarding them therefore leaves the observable side unchanged and weakly lowers the critical value.

Second, we apply *Berk–Jones weighting* (Berk and Jones, 1979). Since the shocks are independent, their count below a threshold is binomial within each cell. This motivates the score $\hat{m}(z_D)\mathrm{KL}(p\|q)$, which accounts for both cell size and the candidate probability, where

$$\mathrm{KL}(p\|q)=p\log(p/q)+(1-p)\log((1-p)/(1-q)).$$

The idea is to measure the discrepancy between two probabilities on a common scale reflecting tail probabilities, making discrepancies comparable across cells. To score only violations of the observable bounds, define

$$d_{\mathrm{BJ}}(a,b;q):=\begin{cases}\mathrm{KL}(a\|q), & q<a,\\ \mathrm{KL}(b\|q), & q>b,\\ 0, & a\le q\le b.\end{cases}$$

The criterion is then

$$\widehat{Q}_n^{\mathrm{BJ}}(\overline{\theta})=\max_{z_D\in\hat{\mathcal{Z}}_D}\ \sup_{c\in C(z_D;\theta,Z)}\hat{m}(z_D)\,d_{\mathrm{BJ}}\Big(\hat{p}_L^{(n)},\hat{p}_U^{(n)};F_\varepsilon(c;\theta_\varepsilon)\Big),$$

with the bounds evaluated at $(z_D,c;\theta)$. The matching control statistic replaces the

observable bounds with the empirical shock frequency:

$$R_n^{\mathrm{BJ}}(\overline{\theta}) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c \in C(z_D;\theta,Z)} \hat{m}(z_D)\, \mathrm{KL}\Big(\hat{F}_n(c; z_D) \Big\| F_\varepsilon(c; \theta_\varepsilon)\Big)\,.$$

At the truth, the empirical shock frequency lies between the bounds, so $\widehat{Q}_n^{\mathrm{BJ}} \leq R_n^{\mathrm{BJ}}$. Simulating the control statistic under the candidate error distribution supplies the critical value, preserving finite-sample coverage by Theorem 4. Supplemental Appendix S.1 provides the implementation details.

4 Simulation Study

This section evaluates the performance of the CSs of Section 3 across network sizes from 100 to 10,000. Throughout, every reported CS is computed from a *single* simulated network. Three further exercises are presented in Supplemental Appendix S.3: common and dyadic shock dependence, multidimensional exogenous covariates, and the symmetry-restricted semiparametric procedure.

4.1 Design

We consider the following model specification

$$Y_{ij} = \mathbf{1}\Big\{\beta_0\,|z_i - z_j| + \gamma_0\,\overline{\mathrm{CF}}_{ij}(Y) - b_0 \;\geq\; \varepsilon_{ij}\Big\}, \tag{18}$$

with true coefficients $\beta_0 = -1$ and $\gamma_0 = 16$, ε_{ij} drawn i.i.d. from Logistic$(0,1)$, z_i drawn i.i.d. uniformly on a grid of 21 equally spaced points in $[-10, 10]$, and $\overline{\mathrm{CF}}_{ij}(Y) := \mathrm{CF}_{ij}(Y)/(n-2)$ being the normalized common-friend count on the equilibrium network Y. This normalization ensures that the data-generating process, and thus the equilibrium network, stays stable across n. Conditioning cells ($z_D \in \mathcal{Z}_D$) are the 231 exact unordered type pairs. We consider network size $n \in \{100, 200, 400, 800, 1{,}600, 3{,}200, 6{,}400, 10{,}000\}$ with $K = 500$ independent replications at every size. We calibrate b_0 via pilot simulation runs, on seeds disjoint from the $K = 500$ production seeds, so that the outcome networks at $n = 400$ have densities around 20%, which yields $b_0 = -0.6$, fixed thereafter.

The choice of the strategic coefficient $\gamma_0 = 16$ seems "large", but its magnitude should be interpreted together with the scale of $X_{ij} := \mathrm{CF}_{ij}(Y)/(n-2)$, which is normalized by $1/(n-2)$. Table 1 reports summary statistics about the equilibrium

Table 1: Summary statistics about the simulated networks

n	density mean (%)	density range (%)	mean X	sd X	mean max X
100	20.40	[12.93, 54.73]	0.054	0.085	0.324
200	20.56	[14.30, 35.55]	0.050	0.075	0.306
400	20.44	[16.17, 31.05]	0.047	0.067	0.281
800	20.21	[17.58, 27.38]	0.044	0.061	0.254
1,600	20.06	[18.26, 23.57]	0.042	0.057	0.228
3,200	19.97	[18.74, 22.01]	0.042	0.055	0.211
6,400	19.95	[18.99, 21.19]	0.041	0.054	0.197
10,000	19.93	[19.18, 20.81]	0.041	0.054	0.190

Notes: $X = \overline{\text{CF}}$. "mean max X" averages each network's maximum across replications.

networks, showing that the endogenous covariate X is generally small in scale, with mean ranging from 0.041 to 0.054 and sd from 0.054 to 0.085 across network sizes. Hence, $\gamma_0 = 16$ only translates to a very moderate amount of variation in the strategic index term. For example, at $n = 200$ the strategic term contributes $\gamma_0 \cdot \text{mean}(X) \approx 0.80$ logit units at the mean and $\gamma_0 \cdot \text{sd}(X) \approx 1.20$ logit units of variation, against a standard logistic shock ε_{ij}. Equivalently, since $X_{ij} = \text{CF}_{ij}/(n-2)$, at $n = 400$ the value $\gamma_0 = 16$ corresponds to an index increment of $\gamma_0/(n-2) \approx 0.04$ per raw common friend, comparable to the per-common-friend effects found in the empirical applications of Section 5.

Since $\overline{\text{CF}}$ is nondecreasing in the network and $\gamma_0 > 0$, the network formation game is supermodular, and the equilibrium networks are computed as the least fixed points by iterations from the empty graph. While our inference procedure requires no supermodularity, no equilibrium selection, and no computation of equilibrium networks, for simulations we do need to generate the simulated equilibrium networks that our inference procedure takes as data. Exploiting the supermodular structure enables us to compute large equilibrium networks even with $n = 10{,}000$.

Throughout this section, we compute various CSs as described in Section 3 and evaluate their performance across the $K = 500$ repetitions, in every case at the 95% nominal level. In the parametric case, the standard logistic CDF fixes the location and scale. We work with a discrete grid on the parameter space: $b_0 - b_{0,\text{true}} \in [-7, 7]$ with step size 0.25, $\beta \in [-3.25, 1.25]$ with step size 0.125, and $\gamma \in [-16, 72]$ with step size 0.125. The test statistic is the Berk–Jones statistic in Section 3.3, with the

Table 2: Parametric confidence sets

n	γ sign	γ width	vac	γ edge L / U	β sign	β width	b_0 width
100	0.748	48.31	0.074	0.106 / 0.176	0.182	2.25	7.00
200	0.948	33.44	0.012	0.026 / 0.024	0.818	1.50	5.00
400	1.000	24.88	0.000	0.000 / 0.002	0.990	0.94	3.50
800	1.000	18.13	0.000	0.000 / 0.000	1.000	0.63	2.50
1,600	1.000	12.63	0.000	0.000 / 0.000	1.000	0.25	1.50
3,200	1.000	7.75	0.000	0.000 / 0.000	1.000	0.13	1.00
6,400	1.000	2.88	0.000	0.000 / 0.000	1.000	0.00	0.00
10,000	1.000	2.25	0.000	0.000 / 0.000	1.000	0.00	0.00

Notes: "Sign" is the share of replications whose projection lies strictly on the correct side of zero; "width" the median total length of the accepted projection, which need not be an interval; "vac" the share accepting the whole γ grid; "edge L / U" the shares touching its lower or upper endpoint.

thresholds c fixed on 731 points spanning $[-36.5, 36.5]$ in steps of 0.1, intersected with the constrained threshold sets $C(z_D; \theta, Z)$ at each candidate. The conditional critical value is simulated accordingly with $B = 999$, and every nonempty type-pair cell is retained. In the semiparametric case, we normalize the scale of the homophily effect parameter to unity, $|\beta| = 1$, fix the location by setting the intercept term b_0 to zero, and retain the cells with at least $\underline{m} = 20$ dyads.

4.2 Parametric Confidence Sets

Table 2 reports the results for the parametric CSs, with the reported diagnostics defined in the notes to the table.

Our inference procedure is demonstrably informative in a single-network setting even at the smallest size $n = 100$. The true parameter vector is covered in every one of the $K = 500$ replications at each tested n, so we omit a coverage column from the table. The CSs exhibit conservative empirical coverage in these simulations, consistent with our theoretical guarantee of at least nominal coverage in finite samples. At the same time, the CSs are nontrivial subsets of the grid, contract with n, and deliver sign-revealing information. The median γ-projection width falls at every tested n, while the vacuous-set rate falls from 7.4% at $n = 100$ to zero from $n = 400$ onward.

Notably, the sign of the strategic coefficient γ, an important object of interest

in the strategic network formation literature (e.g., de Paula and Tang, 2012; Sheng, 2020), is certified in 74.8% of replications at $n = 100$, in 94.8% at $n = 200$, and in every replication from $n = 400$ onward, while the median γ projection width falls from 48.31 at $n = 100$ to 2.25 at $n = 10{,}000$, on a candidate grid of length 88. To the best of our knowledge, single-network finite-sample valid and sign-revealing CSs were not previously available in this environment.

The CSs also deliver increasingly tight bounds on the homophily coefficient β and the intercept b_0. The negative sign of β is certified in 18.2% of replications at $n = 100$, 81.8% at $n = 200$, 99.0% at $n = 400$, and in every replication from $n = 800$ onward. The median width falls from 2.25 for β and 7.00 for b_0, half of the respective candidate grids, to zero at $n = 6{,}400$ for both parameters. Thus the nuisance-parameter projections concentrate considerably faster than the projection for the strategic coefficient.

4.3 Semiparametric Confidence Sets

In Table 3, we report results on the semiparametric versions of our inference procedure. The semiparametric column is calibrated by conditional simulation in the cellwise form introduced in Footnote 13 and explained further in Supplemental Appendix S.1. We also include corresponding results about the parametric case from the last subsection for ease of comparison.

These CSs are computed on the same simulated networks and use the same $\gamma \in [-16, 72]$ grid with step size 0.125. The parametric procedure treats (b_0, β, γ) as unknown under the standard-logistic location and scale normalization. The fully semiparametric procedure instead fixes $b_0 = 0$ as a location normalization and sets $|\beta| = 1$ as a scale normalization. For each sign of β, we use a threshold grid with step size 0.0625 and a range chosen to cover all candidate index values. Width is the median total length of the accepted γ projection in both columns. Vacuity refers to the γ projection in both columns.

The semiparametric CSs, while less tight than their parametric counterparts as anticipated, remain nontrivial and informative. The true γ_0 is covered in every replication, so the coverage column is similarly omitted as in Table 2. The cleanest comparison between the two CSs concerns the *sign rate* for γ_0, which orders the CSs by the strength of the imposed assumptions. Despite leaving F_ε unrestricted, the semi-

Table 3: Semiparametric confidence sets: γ projections

n	parametric			semiparametric		
	sign	width	vac	sign	width	vac
100	0.748	48.31	0.074	0.416	72.00	0.094
200	0.948	33.44	0.012	0.560	67.63	0.018
400	1.000	24.88	0.000	0.666	53.94	0.000
800	1.000	18.13	0.000	0.716	40.00	0.000
1,600	1.000	12.63	0.000	0.868	23.38	0.000
3,200	1.000	7.75	0.000	0.978	20.00	0.000
6,400	1.000	2.88	0.000	0.998	16.25	0.000
10,000	1.000	2.25	0.000	0.998	12.50	0.000

parametric CSs increasingly reveal the positive sign of γ_0: the sign rate rises from 41.6% at $n = 100$ to 97.8% at $n = 3{,}200$, reaching 99.8% (499 of 500 replications) at both $n = 6{,}400$ and $n = 10{,}000$. By comparison, the parametric CSs certify the sign in every replication from $n = 400$ onward. The semiparametric projections remain wider but contract substantially, with median width falling from 72.00 at $n = 100$ to 12.50 at $n = 10{,}000$. Although boundary contact may understate their widths at small n, the projections contract within the candidate range at larger sizes. The vacuity rate falls from 9.4% at $n = 100$ to 1.8% at $n = 200$, with no vacuous projections from $n = 400$ onward.

The Supplemental Appendix examines shock dependence, multidimensional exogenous covariates, and symmetry restrictions on F_ε. Additive common shocks leave performance largely unchanged, whereas block dependence outside the maintained assumptions leads to undercoverage. Additional covariates preserve informative sign results at larger network sizes, while imposing symmetry improves sign revelation relative to the unrestricted semiparametric procedure.

5 Empirical Applications

We apply the parametric CS to two empirical data sets: contact (and friendship) among high-school students, and friendship among Twitch streamers. In each application we ask whether link formation responds to the links of others, by constructing a finite-sample valid CS for the coefficient on a common-friend statistic while leaving

Table 4: High-school network: summary statistics

network	links	density (%)	mean degree	CF mean (sd)	CF ≥ 1 (%)
contact, $\underline{\kappa} = 1$	5,818	10.92	35.58	4.33 (6.64)	73.20
contact, $\underline{\kappa} = 2$	4,031	7.56	24.65	2.11 (4.62)	40.91
contact, $\underline{\kappa} = 3$	3,337	6.26	20.41	1.46 (3.59)	32.11
friendship	406	0.76	2.48	0.05 (0.43)	2.47

equilibrium selection and network dependence unrestricted. In both data sets that coefficient is bounded away from zero from below at the 95% confidence level, so triad closure is an economically operative force in these networks and not an artifact of homophily on the observed covariates. This distinction matters for policy: interventions that seed or remove a small number of links have knock-on effects on the remaining ones precisely when the strategic channel is present, whereas under pure homophily their effects would remain local. In both applications, each reported row is the projection of a single joint 95% confidence set over all coefficients.

5.1 High-School Contact and Friendship Networks

The Thiers13 data contain 327 students, and hence 53,301 dyads, in nine classes and four class types (Mastrandrea et al., 2015).[15] Contact sensors record face-to-face proximity in twenty-second intervals during one school week, and we set $Y_{ij} = 1$ when a pair has at least $\underline{\kappa}$ recorded intervals in total. The baseline is $\underline{\kappa} = 2$, with $\underline{\kappa} = 1$ and $\underline{\kappa} = 3$ also reported as robustness checks. The friendship layer sets $Y_{ij} = 1$ if either student nominates the other in a survey.[16] The exogenous covariates are an indicator for different classes (D_{ij}) and an indicator for different class types (T_{ij}), so $Z_{ij} = (1, D_{ij}, T_{ij})$, and the endogenous covariate is the common-friend count in the same network, normalized as in the simulations and reported below in per-common-friend units. Selected summary statistics are reported in Table 4.

Again, we use the Berk–Jones statistic with constrained thresholding in Section 3.3. The conditioning cells are the 45 unordered pairs of the nine classes, with a

[15]The data set is publicly available and can be accessed at https://sociopatterns.org/datasets/high-school-contact-and-friendship-networks/.

[16]Only 135 of the 327 students returned the friendship survey, so the friendship layer is a partially observed version of the underlying nomination network. We therefore treat the contact networks as the primary specification and report the friendship results as supporting evidence.

Table 5: High-school network: parametric confidence-set projections

network	b_D	b_T	γ
contact, $\underline{\kappa} = 1$	$[-4.250, 2.750]$	$[-4.000, 2.500]$	$[0.020, 0.335]$
contact, $\underline{\kappa} = 2$	$[-4.500, 1.250]$	$[-3.500, 1.000]$	$[0.045, 0.370]$
contact, $\underline{\kappa} = 3$	$[-4.750, 0.750]$	$[-3.500, 0.500]$	$[0.025, 0.405]$
friendship	$[-4.500, 2.750]$	$[-5.250, 3.250]$	$[0.110, 2.270]$

cell floor of 20 dyads. The critical values use $B = 999$ conditional simulation draws. The endogenous statistic is $X = \mathrm{CF}/(n-2)$. We use the normalized common-friend count $X = \mathrm{CF}/(n-2)$, with grid spacing 1.625 for its coefficient γ. For ease of interpretation, we report $\gamma/(n-2)$, the increase in the link index per additional common friend. Since $n - 2 = 325$, the reported grid spacing is 0.005. We search over a rectangular parameter grid, with spacing 0.25 for the remaining coefficients. All reported confidence-set projections lie strictly inside their respective search ranges, so the boundaries are nonbinding. The threshold grid has spacing 0.125 and covers the full range of the link index in each specification.

Table 5 reports the parametric 95%-level confidence-set projection results under the assumption that ε_{ij} is standard logistic. All four common-friend projections are positive. Equivalently, the $\gamma = 0$ slice of each reported joint CS, which is the CS for the corresponding Z-only logistic specification under the same inversion, is empty, so by Remark 3 the specification without the common-friend term is rejected at the 5% level at every nuisance-coefficient value on the reported grid. A positive γ indicates the tendency for triad closure in strategic network formation: sharing a contact (or friend) raises the net surplus from linking. In the baseline $\underline{\kappa} = 2$ contact network one additional common friend raises the structural link index by 0.045 to 0.370. At the sample mean of 2.11 common friends, the endogenous term contributes roughly 0.09 to 0.78 index units, relative to the standard logistic error scale. The $\underline{\kappa} = 1$ and $\underline{\kappa} = 3$ rows are also strictly positive, so the sign of γ does not depend on a single contact threshold. The friendship projection for γ_0 is also positive but much wider than those for the contact networks. Partial survey response (Footnote 16) limits identifying variation: 97.53% of dyads have no common friend, and 193 of 327 students are isolated.

Table 6: Twitch gamer network: summary statistics

language	nodes	links	density (%)	CF mean (sd)	CF ≥ 1 (%)
German (DE)	9,498	153,138	0.34	0.86 (2.51)	37.21
English (EN)	7,126	35,324	0.14	0.08 (0.41)	6.31
Spanish (ES)	4,648	59,382	0.55	0.66 (2.14)	27.72
French (FR)	6,549	112,666	0.53	1.09 (2.95)	38.94
Portuguese (PT)	1,912	31,299	1.71	2.18 (5.06)	51.32
Russian (RU)	4,385	37,304	0.39	0.46 (1.52)	23.25

5.2 Twitch Friendship Networks

The Twitch data set contains information about networks of gamers who stream in a certain language (DE, EN, ES, FR, PT, and RU): six undirected friendship networks, one per streamer language (Rozemberczki et al., 2021).[17] We treat each network as a separate inference problem with language-specific parameters, and the CS is computed separately for each network. See Table 6 for some key summary statistics.

We construct the exogenous covariates as follows. Let $q_i^V \in \{1, \ldots, 5\}$ be streamer i's within-language quintile of total views and m_i the mature-content indicator. Define

$$V_{ij} = |q_i^V - q_j^V|, \qquad S_{ij} = q_i^V + q_j^V - 6, \qquad M_{ij} = m_i + m_j - 1,$$

so that $V_{ij} \in \{0, \ldots, 4\}$ measures the popularity gap, $S_{ij} \in \{-4, \ldots, 4\}$ captures the popularity level of the pair, and $M_{ij} \in \{-1, 0, 1\}$ the centered mature-content pair count. All three are integer-valued and symmetric in (i, j), and together they give the 15 unordered views-quintile pairs crossed with the three mature-pair values. The structural index in the language-ℓ network is then

$$\delta_{ij,\ell} = b_{0\ell} + b_{V\ell} V_{ij} + b_{S\ell} S_{ij} + b_{M\ell} M_{ij} + \gamma_\ell \, X_{ij,\ell},$$

with $Z_{ij} = (1, V_{ij}, S_{ij}, M_{ij})$ and the endogenous regressor X_{ij} given by

$$X_{ij,\ell} := \min \left\{ \mathrm{CF}_{ij}, \, q_{95,\ell}^{\mathrm{CF}} \right\},$$

where $q_{95,\ell}^{\mathrm{CF}}$ is the 95th percentile of the raw common-friend counts in network ℓ (the q_{95} column of Table 7): $q_{95,\ell}^{\mathrm{CF}} = 4, 1, 3, 5, 9$, and 2 for DE, EN, ES, FR, PT, and RU,

[17] The data set is publicly available and can be accessed at https://snap.stanford.edu/data/twitch-social-networks.html.

Table 7: Twitch gamer network: right tails of CF distribution

language	CF = 0 (%)	q_{90}	q_{95}	q_{99}	$q_{99.9}$	max CF
German (DE)	62.79	2	4	9	24	1,432
English (EN)	93.69	0	1	2	4	134
Spanish (ES)	72.28	2	3	8	24	436
French (FR)	61.06	3	5	11	26	1,095
Portuguese (PT)	48.68	6	9	20	54	449
Russian (RU)	76.75	1	2	6	14	562

respectively. Since the cap is computed from the realized network, $X_{ij,\ell}$ is itself a globally defined endogenous statistic, which our framework explicitly accommodates as discussed in Footnote 4.

We cap the common-friend count at the network's own tail quantile to deal with the highly skewed and heterogeneous distribution of CF_{ij} within each language network and across the six language networks. Table 7 documents how extreme that skewness is. In the German network, for instance, 62.79% of dyads have no common friend at all and the 99th percentile is 9, yet the maximum CF_{ij} is 1,432. An uncapped count would therefore let extreme neighborhoods dominate identifying variation. The transformation $\log(1 + \mathrm{CF}_{ij})$ accommodates zeros but imposes diminishing returns. Capping limits the upper tail while preserving zeros and constant per-friend index increments below the cap. All six networks are heavily skewed, but the tails differ substantially across the six: the 95th percentile ranges from 1 (English) to 9 (Portuguese), so a common cap would represent very different events across networks. Capping at the network's own 95th percentile instead preserves the dyad-level variation that identifies γ_ℓ while adapting to each network's tail: under this specification γ_ℓ is the change in the latent link index per additional common friend below the cap.

We use fine threshold and parameter grids with steps of 0.125 and 0.025, respectively. We use sufficiently wide parameter ranges so that no reported projection reaches a grid boundary. The conditioning cells are the 45 combinations of the 15 views-quintile pairs and the three values of M_{ij}, subject to a 20-dyad cell floor which no cell reaches. The critical values use $B = 999$ conditional simulation draws.

Table 8 reports the parametric confidence-set projections under the assumption that ε_{ij} is standard logistic. Every language produces a strictly positive CS for γ_ℓ: under the maintained capped-index specification, each additional common friend below

Table 8: Twitch gamer network: parametric confidence-set projections

language	b_V (views gap)	b_S (views level)	b_M (mature pair)	γ (capped CF)
German (DE)	$[0.025, 0.600]$	$[0.175, 0.725]$	$[-0.650, 0.425]$	$[0.350, 2.250]$
English (EN)	$[0.100, 0.425]$	$[0.350, 0.600]$	$[-0.275, 0.225]$	$[0.650, 8.475]$
Spanish (ES)	$[-0.050, 0.725]$	$[0.125, 0.750]$	$[-1.175, 0.800]$	$[0.325, 2.850]$
French (FR)	$[-0.275, 0.800]$	$[0.050, 0.875]$	$[-0.900, 0.900]$	$[0.225, 1.825]$
Portuguese (PT)	$[-0.250, 1.125]$	$[0.000, 1.100]$	$[-1.225, 1.175]$	$[0.100, 1.000]$
Russian (RU)	$[0.175, 0.750]$	$[0.275, 0.800]$	$[-0.600, 0.600]$	$[0.400, 4.100]$

the cap raises the latent link index. The exclusion of zero also has a restricted-model interpretation. When $\gamma_\ell = 0$, $X_{ij,\ell}$ drops out of the index, so the zero slice of the joint inversion is exactly the CS for the Z-only logistic specification, computed with the same cells, threshold grid, and critical value. This slice is empty in all six language networks: by the emptiness argument of Remark 3, the Z-only specification is rejected in each network over the stated coefficient ranges at the 5% level. These results highlight the importance of accounting for endogenous network variables, such as common-friend counts, in this application.

The per-friend lower endpoints range from 0.100 (Portuguese) to 0.650 (English). Under the standard-logistic normalization, the German projection $[0.350, 2.250]$, for instance, implies a latent-index increase of 0.350 to 2.250 per additional common friend below the cap, while the English lower bound is an increase of 0.650 from the first common friend. Because common friends are endogenous, these are not conditional log-odds or odds-ratio effects. The exogenous-coefficient projections also provide sign information: the views-level projections are strictly positive in five networks, and the views-gap projections are strictly positive in the German, English, and Russian networks. The mature-content projections include zero in all six networks. Across networks with very different sizes, densities, and tail behaviors, the procedure delivers informative bounds on the strategic coefficient and reveals the signs of several effects associated with observed streamer characteristics.

6 Conclusion

This paper develops single-network and finite-sample valid inference for structural parameters in strategic network formation models. The bounding-by-c device converts

restrictions involving endogenous network statistics into realization-wise inequalities whose middle term depends only on exogenous covariates and i.i.d. pairwise shocks. This yields simulable critical values and finite-sample CSs in both parametric and semiparametric settings, without the need to solve, simulate, or enumerate equilibrium networks. The validity of the procedure does not require restrictions on network density, network dependence, or equilibrium selection. Its computational simplicity also makes the procedure scalable to large networks.

The core idea of bounding by c might be useful in other contexts as well, especially in settings with complicated and intractable endogeneity. Potential applications include nontransferable-utility network models with bilateral consent, as well as matching, auctions, and entry games. Developing these model-specific bounds and assessing their informativeness are left for future research.

Appendix: Proofs

Proof of Lemma 1. Fix a dyad type z_D, write $N_n := \binom{n}{2}$ for the number of dyads, and define, for $c \in \mathbb{R}$,

$$\hat{A}_{n,z_D}(c) := \frac{1}{N_n} \sum_{ij} \mathbf{1}\{\varepsilon_{ij} \le c\} \mathbf{1}\{Z_{D,ij} = z_D\},$$

$$\hat{A}^{-}_{n,z_D}(c) := \frac{1}{N_n} \sum_{ij} \mathbf{1}\{\varepsilon_{ij} < c\} \mathbf{1}\{Z_{D,ij} = z_D\},$$

$$\hat{D}_{n,z_D} := N_n^{-1} \sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\} \equiv N_n^{-1} \hat{m}\left(z_D\right),$$

and $\hat{F}_n(c; z_D) = \hat{A}_{n,z_D}(c)/\hat{D}_{n,z_D}$, $\hat{F}^{-}_n(c; z_D) := \hat{A}^{-}_{n,z_D}(c)/\hat{D}_{n,z_D}$ whenever $\hat{D}_{n,z_D} > 0$.

We first establish pointwise convergence results at each z_D and each c. Each of $\hat{A}_{n,z_D}(c)$, $\hat{A}^{-}_{n,z_D}(c)$, and $\hat{D}_{n,z_D}$ is an average over the N_n dyads of a bounded kernel of the array $W_{ij} := (Z_i, Z_j, \varepsilon_{ij})$. Under Assumptions 2–4 each of the three scalar summand arrays (for fixed c and z_D) is symmetric in (i, j) by the symmetry of the partition cells, jointly exchangeable (i.e., its distribution is invariant under relabeling of the agents), and dissociated (entries indexed by disjoint agent sets are independent). The strong law of large numbers for U-statistics of such arrays (e.g., Theorem 3 of Eagleson and Weber, 1978) therefore gives, almost surely,

$$\hat{A}_{n,z_D}(c) \to F_\varepsilon(c)\, q(z_D), \quad \hat{A}^{-}_{n,z_D}(c) \to F_\varepsilon(c^-)\, q(z_D), \quad \hat{D}_{n,z_D} \to q(z_D) > 0,$$

where $q(z_D) := \mathbb{P}(Z_{D,ij} = z_D)$, $F_\varepsilon(c^-) := \lim_{u \uparrow c} F_\varepsilon(u)$ and the products on the right-hand side, e.g., $\mathbb{E}[\mathbf{1}\{\varepsilon_{ij} \le c\}\mathbf{1}\{Z_{D,ij} = z_D\}] = F_\varepsilon(c)q(z_D)$, follow from the independence between ε and Z (Assumption 3). Hence, $\hat{F}_n(c; z_D) \to F_\varepsilon(c)$ and $\hat{F}_n^-(c; z_D) \to F_\varepsilon(c^-)$ for each fixed c, almost surely.

Next, we establish uniformity in c at each z_D. Fix an integer $M \ge 2$ and let $c_j := \inf\{c : F_\varepsilon(c) \ge j/M\}$, $j = 1, \ldots, M-1$, denote the M-quantile grid of F_ε; each c_j is finite because F_ε tends to 0 and 1 at $\mp\infty$. Let Ω_{M,z_D} be the probability-one event on which $\hat{D}_{n,z_D} > 0$ eventually and the pointwise convergence established above holds for both $\hat{F}_n$ and $\hat{F}_n^-$ at every c_j. Define

$$\Delta_{n,z_D,M} := \max_{1 \le j \le M-1} \max\left\{ \left| \hat{F}_n(c_j; z_D) - F_\varepsilon(c_j) \right|, \ \left| \hat{F}_n^-(c_j; z_D) - F_\varepsilon(c_j^-) \right| \right\}.$$

Since the grid is finite, $\Delta_{n,z_D,M} \to 0$ on Ω_{M,z_D}. To extend this control from the grid to every $c \in \mathbb{R}$, fix c. For the lower bound on $\hat{F}_n(c; z_D)$, if $F_\varepsilon(c) < 1/M$ then $F_\varepsilon(c) - \hat{F}_n(c; z_D) < 1/M$ immediately. Otherwise, set $j = \min\{\lfloor MF_\varepsilon(c) \rfloor, M-1\}$. Then $c_j \le c$ and $F_\varepsilon(c_j) \ge j/M \ge F_\varepsilon(c) - 1/M$, so

$$\hat{F}_n(c; z_D) \ge \hat{F}_n(c_j; z_D) \ge F_\varepsilon(c) - \frac{1}{M} - \Delta_{n,z_D,M}.$$

For the upper bound, if $F_\varepsilon(c) \ge (M-1)/M$ then $\hat{F}_n(c; z_D) - F_\varepsilon(c) \le 1/M$. Otherwise, set $j = \lfloor MF_\varepsilon(c) \rfloor + 1 \le M-1$. Because $j/M > F_\varepsilon(c)$, the definition of c_j implies $c < c_j$ and $F_\varepsilon(c_j^-) \le j/M$. Hence, we have

$$\hat{F}_n(c; z_D) \le \hat{F}_n^-(c_j; z_D) \le F_\varepsilon(c) + \frac{1}{M} + \Delta_{n,z_D,M}.$$

Combining the two bounds gives $\sup_{c \in \mathbb{R}} \left| \hat{F}_n(c; z_D) - F_\varepsilon(c) \right| \le \Delta_{n,z_D,M} + \frac{1}{M}$ on Ω_{M,z_D}.

Finally, $\mathcal{Z}_D$ is finite with $q(z_D) > 0$ for any z_D by Definition 1, so the intersection $\overline{\Omega}_M$ of the finitely many Ω_{M,z_D} still has probability one, and the bound above holds on it with $\Delta_{n,M} := \max_{z_D \in \mathcal{Z}_D} \Delta_{n,z_D,M} \to 0$. Since M is arbitrary and $\bigcap_{M \ge 2} \overline{\Omega}_M$ is again a probability-one event, $\max_{z_D \in \mathcal{Z}_D} \sup_{c \in \mathbb{R}} \left| \hat{F}_n(c; z_D) - F_\varepsilon(c) \right| \to 0$ as $n \to \infty$. □

Proof of Theorem 1. Taking $\limsup_n$ on the left of (9) and applying Lemma 1,

$$p_L^\infty(z_D, c; \theta_0) = \limsup_n \ \hat{p}_L^{(n)}(z_D, c; \theta_0) \ \le \ \limsup_n \ \hat{F}_n(c; z_D) \ = \ F_\varepsilon(c),$$

and symmetrically $p_U^\infty(z_D, c; \theta_0) = \liminf_n \hat{p}_U^{(n)}(z_D, c; \theta_0) \ge \liminf_n \hat{F}_n(c; z_D) = F_\varepsilon(c)$, which together imply that $p_L^\infty(z_D, c; \theta_0) \le F_\varepsilon(c) \le p_U^\infty(z_D, c; \theta_0)$ for all $c \in \mathbb{R}$ and for

all z_D. Then $\max_{z_D} p_L^\infty(z_D, c; \theta_0) \le F_\varepsilon(c) \le \min_{z_D} p_U^\infty(z_D, c; \theta_0)$, giving part (i). Part (ii) follows by inserting $F_\varepsilon(\cdot\,; \theta_{\varepsilon,0}) = F_\varepsilon$ as the middle term. □

Proof of Lemma 2. Fix any $c \in \mathbb{R}$. The finite-sample sandwich (9) holds cell by cell at θ_0: $\hat{p}_L^{(n)}(z_D, c; \theta_0) \le \hat{F}_n(c; z_D) \le \hat{p}_U^{(n)}(z_D, c; \theta_0)$ for every $z_D \in \hat{\mathcal{Z}}_D$. Maximizing the left inequality and minimizing the right one over $z_D \in \hat{\mathcal{Z}}_D$ yields

$$\max_{z_D} \hat{p}_L^{(n)}(z_D, c; \theta_0) \;\le\; \max_{z_D} \hat{F}_n(c; z_D), \quad \min_{z_D} \hat{p}_U^{(n)}(z_D, c; \theta_0) \;\ge\; \min_{z_D} \hat{F}_n(c; z_D),$$

and hence

$$\max_{z_D} \hat{p}_L^{(n)}(z_D, c; \theta_0) - \min_{z_D} \hat{p}_U^{(n)}(z_D, c; \theta_0) \;\le\; \max_{z_D} \hat{F}_n(c; z_D) - \min_{z_D} \hat{F}_n(c; z_D).$$

The right-hand side is at most R_n by the definition (13) and is nonnegative. Hence, taking the supremum over c and applying the positive part function preserves the bound, implying $\widehat{Q}_n(\theta_0) \le R_n$. □

Proof of Theorem 2. We condition on a fixed realization of $Z = (Z_i)_{i=1}^n$, and suppress the qualifier "almost surely" subsequently. Then, $\hat{\mathcal{Z}}_D$ and the cell sizes $\hat{m}(z_D)$ are all fixed. Let $(\nu_{ij})_{ij}$ be i.i.d. Unif$[0,1]$ random variables, independent of everything else, and define the *distributional transform*

$$U_{ij} := F_\varepsilon(\varepsilon_{ij}^-) + \nu_{ij}\big[F_\varepsilon(\varepsilon_{ij}) - F_\varepsilon(\varepsilon_{ij}^-)\big].$$

When F_ε is continuous, the above specializes to the standard probability integral transform $U_{ij} := F_\varepsilon(\varepsilon_{ij})$. Regardless of whether F_ε is continuous or not, it is a standard result that $U_{ij} \sim$ Unif$[0,1]$ and $\varepsilon_{ij} = F_\varepsilon^{-1}(U_{ij})$, where $F_\varepsilon^{-1}(u) := \inf\{x : F_\varepsilon(x) \ge u\}$ is the generalized inverse of F_ε. Using the fact that $F_\varepsilon^{-1}(u) \le c \iff u \le F_\varepsilon(c)$, we have $\mathbf{1}\{\varepsilon_{ij} \le c\} = \mathbf{1}\{U_{ij} \le F_\varepsilon(c)\}$ and hence

$$\hat{F}_n(c; z_D) = G_{z_D}\big(F_\varepsilon(c)\big),$$

where G_{z_D} is defined as the empirical distribution function of $\hat{m}(z_D)$ i.i.d. uniform random variables at cell z_D. The independence follows from the fact that each cell contains different dyads across which ε_{ij} is assumed to be independent (Assumption 2). Hence, G_{z_D} has exactly the same conditional distribution as the simulated $G_{z_D}^{(b)}$. Substituting $\hat{F}_n(c; z_D) = G_{z_D}\big(F_\varepsilon(c)\big)$ into (13) and writing $\mathcal{R} := F_\varepsilon(\mathbb{R}) \subseteq [0,1]$ for

the range of F_ε, we have

$$R_n = \sup_{u\in\mathcal{R}}\Big[\max_{z_D\in\hat{\mathcal{Z}}_D} G_{z_D}(u) - \min_{z_D\in\hat{\mathcal{Z}}_D} G_{z_D}(u)\Big] \\ \le \sup_{u\in[0,1]}\Big[\max_{z_D\in\hat{\mathcal{Z}}_D} G_{z_D}(u) - \min_{z_D\in\hat{\mathcal{Z}}_D} G_{z_D}(u)\Big] =: R^*.$$

The simulated $R^{(b)}$ then shares exactly the same distribution as R^*. If F_ε is continuous, then $R_n = R^*$. When F_ε has atoms, $R_n \neq R^*$ in general, but $R_n \le R^*$ is always true and thus R^* remains a valid device to control size as shown below. Since $(R^*, R^{(1)}, \ldots, R^{(B)})$ is conditionally i.i.d. given Z, it is exchangeable, and thus the ranking of R^* in $(R^*, R^{(1)}, \ldots, R^{(B)})$ is uniform over $\{1, \ldots, B+1\}$ up to ties. Therefore, the finite-B-adjusted order statistic satisfies

$$\mathbb{P}\big(R^* > \hat{q}^R_{n,1-\alpha}(Z) \,\big|\, Z\big) \;\le\; \frac{B+1-\lceil(1-\alpha)(B+1)\rceil}{B+1} \;\le\; \alpha,$$

where ties can only reduce the rejection probability. Since $\widehat{Q}_n(\theta_0) \le R_n$ by Lemma 2 and $R_n \le R^*$ as shown above, we conclude that

$$\mathbb{P}\big(\widehat{Q}_n(\theta_0) > \hat{q}^R_{n,1-\alpha}(Z) \,\big|\, Z\big) \le \mathbb{P}\big(R^* > \hat{q}^R_{n,1-\alpha}(Z) \,\big|\, Z\big) \le \alpha.$$

□

Proof of Lemma 3. Condition on Z. Fix a constant $c \in \mathbb{R}$ as before. Write $F_\varepsilon(c) = F_\varepsilon(c; \theta_{\varepsilon,0})$ in short for the true CDF. By the sandwich inequality (9) and the definition of $R_n^{\text{par}}(\theta_{\varepsilon,0})$, for every $z_D \in \hat{\mathcal{Z}}_D$,

$$\hat{p}_L^{(n)}(z_D, c; \theta_0) - F_\varepsilon(c) \;\le\; \hat{F}_n(c; z_D) - F_\varepsilon(c) \;\le\; R_n^{\text{par}}(\theta_{\varepsilon,0}),$$
$$F_\varepsilon(c) - \hat{p}_U^{(n)}(z_D, c; \theta_0) \;\le\; F_\varepsilon(c) - \hat{F}_n(c; z_D) \;\le\; R_n^{\text{par}}(\theta_{\varepsilon,0}).$$

Taking maxima over z_D and suprema over c as before yields the result. □

Proof of Theorem 3. Condition on Z. As before, by Assumptions 2 and 3, the real shock array $\{\varepsilon_{ij}\}_{ij}$ and the B simulated arrays $\{\varepsilon_{ij}^{(b)}\}_{ij}$ are i.i.d. draws from the same distribution given Z, so the realized statistic $R_n^{\text{par}}(\theta_{\varepsilon,0})$ and the simulated statistics $R_n^{\text{par},(1)}(\theta_{\varepsilon,0}), \ldots, R_n^{\text{par},(B)}(\theta_{\varepsilon,0})$ are conditionally i.i.d. and hence exchangeable. Combined with Lemma 3, the finite-B rank argument proceeds in the same way as in the proof of Theorem 2. □

Proof of Theorem 4. At the true parameter $\overline{\theta}_0$, (9) gives $V^{\pm}(z_D, c; \overline{\theta}_0) \le$

$R^{\pm}(z_D, c; \theta_{\varepsilon,0})$ for every coordinate and every realization, so monotonicity of $T_{Z,\bar{\theta}_0}$ in each argument yields $T_{Z,\bar{\theta}_0}\big(V(\overline{\theta}_0)\big) \leq T_{Z,\bar{\theta}_0}\big(R(\theta_{\varepsilon,0})\big)$. Conditional on Z, the realized $T_{Z,\bar{\theta}_0}(R)$ and its simulated counterparts $T_{Z,\bar{\theta}_0}(R^{(1)}), \ldots, T_{Z,\bar{\theta}_0}(R^{(B)})$ are again i.i.d., each distributed as $T_{Z,\bar{\theta}_0}\big(R(\theta_{\varepsilon,0})\big)$. The rest of the proof is exactly the same as that of Theorem 2. □

Supplemental Appendix to Single-Network Finite-Sample Inference in Strategic Network Formation Models

Wayne Yuan Gao and Ming Li

September 23, 2026

S.1 Constrained Thresholding and Berk–Jones Weighting

This section supplies the details of the improved parametric procedure of Section 3.3 of the main paper. Its two ingredients, a restriction of the thresholds to the informative window $C(z_D; \theta, Z)$ and a reweighting of the surviving inequalities on the Berk–Jones scale, are best seen as two applications of one principle. Theorem 4 of the main paper states that any known aggregation $T_{Z,\overline{\theta}}$ of the one-sided violations $V^{\pm}(z_D, c; \overline{\theta})$ that is nondecreasing in each coordinate delivers a valid confidence set (CS), provided the same aggregation is applied to the simulated control coordinates $R^{\pm}(z_D, c; \theta_\varepsilon)$. Constrained thresholding is the aggregation that sets to zero the coordinates at which the observable envelopes cannot respond to c. Berk–Jones weighting is the aggregation that rescales each remaining coordinate by its cell size, so that violations are measured in units of null tail probability rather than of raw frequency. We describe the two aggregations in turn, verify that their combination is nondecreasing, explain which part of the construction carries over to the semiparametric procedure, and record the cellwise weighting used in the semiparametric columns of our simulations. We retain θ for the structural candidate and $\overline{\theta} = (\theta', \theta_\varepsilon')'$ for the full parametric candidate. In the parametric calculations below, $F_\varepsilon(c)$ abbreviates $F_\varepsilon(c; \theta_\varepsilon)$ at a fixed candidate error-distribution parameter θ_ε.

S.1.1 Constrained Threshold Sets

Recall from model (1) that the function w_n generates the dyadic covariate $Z_{ij} = w_n(Z_i, Z_j)$ from the individual covariates. The notation "$\tilde{z}_D = z_D$" means that the discretized version $\tilde{z}_D$ of an individual-covariate vector $\tilde{z}$, in the sense of Definition 1, falls in the cell z_D. The constrained threshold set is

$$C(z_D; \theta, Z) = \Big[\inf_{\tilde{z}:\, \tilde{z}_D = z_D} w_n(\tilde{z})' \beta + \inf_{x \in \mathcal{X}} \gamma' x, \quad \sup_{\tilde{z}:\, \tilde{z}_D = z_D} w_n(\tilde{z})' \beta + \sup_{x \in \mathcal{X}} \gamma' x \Big].$$

Here $\mathcal{X}$ is a known deterministic set containing every value the statistic X_{ij} can take, for example $\mathcal{X} = \{0, 1, \ldots, n-2\}$ for a raw common-friend count and $\mathcal{X} = [0, 1]$ for its normalized version. The realized X is never used in constructing the threshold sets, which keeps them (θ, Z)-measurable. Coverage does not depend on $\mathcal{X}$ being correct, since any (θ, Z)-measurable window yields a valid CS by the monotonicity argument below; a set $\mathcal{X}$ that is too small affects only the informativeness of the CS.

Thresholds outside $C(z_D; \theta, Z)$ carry no information about θ. Below the window, every dyad in the cell has $\delta_{ij}(\theta) > c$, so $\hat{p}_L^{(n)} = 0$ and $\hat{p}_U^{(n)} = \overline{Y}(z_D)$, the link frequency of the cell. Above it, every dyad has $\delta_{ij}(\theta) < c$, so $\hat{p}_L^{(n)} = \overline{Y}(z_D)$ and $\hat{p}_U^{(n)} = 1$. Neither envelope varies with c in either region, while $F_\varepsilon(c)$ is nondecreasing, so the largest violation attainable below the window occurs at its lower endpoint and the largest violation attainable above the window occurs at its upper endpoint. Both endpoints belong to the closed window, so restricting the supremum over c to $C(z_D; \theta, Z)$ leaves the observable side of the test unchanged. The control coordinates $R^{\pm}(z_D, c; \theta_\varepsilon)$, which involve only the simulated shocks and not the index, would nonetheless keep contributing to the simulated maximum outside the window and thus inflate the critical value. Discarding those coordinates therefore weakly lowers the critical value at no cost on the observable side, and the improvement is strict whenever the simulated maxima are attained outside the constrained sets sufficiently often. When $\mathcal{X}$ is compact, as for a normalized statistic, the restriction can be substantial.

If $\mathcal{X}$ is unbounded but sign-restricted, the same formula still trims away unnecessary thresholds on one side. For scalar X_{ij} with $\mathcal{X} = [0, \infty)$,

$$C(z_D; \theta, Z) = \begin{cases} \big[\inf_{\tilde{z}:\, \tilde{z}_D = z_D} w_n(\tilde{z})' \beta, \ +\infty \big), & \gamma > 0, \\ \big[\inf_{\tilde{z}:\, \tilde{z}_D = z_D} w_n(\tilde{z})' \beta, \ \sup_{\tilde{z}:\, \tilde{z}_D = z_D} w_n(\tilde{z})' \beta \big], & \gamma = 0, \\ \big(-\infty, \ \sup_{\tilde{z}:\, \tilde{z}_D = z_D} w_n(\tilde{z})' \beta \big], & \gamma < 0. \end{cases}$$

Constrained threshold sets need not be combined with the Berk–Jones weights. Used with the plain maximum, they induce the test statistic and the control statistic

$$\widehat{Q}_n^{\mathrm{con}}(\overline{\theta}) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c \in C(z_D;\theta,Z)} \Big[\max\Big\{ \hat{p}_L^{(n)}(z_D, c; \theta) - F_\varepsilon(c),\ F_\varepsilon(c) - \hat{p}_U^{(n)}(z_D, c; \theta) \Big\} \Big]_+,$$

$$T_{\overline{\theta}}^{\mathrm{con}}(R^{(b)}) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c \in C(z_D;\theta,Z)} \big| \hat{F}_n^{(b)}(c; z_D) - F_\varepsilon(c) \big|,$$

where $[x]_+ := \max\{x, 0\}$ as in the main text. These are the statistics of Section 3.2 of the main paper with the coordinates outside the windows removed, and the CS obtained by inverting the test against the $\lceil (1-\alpha)(B+1) \rceil$-th smallest value of $\big(T_{\overline{\theta}}^{\mathrm{con}}(R^{(b)})\big)_{b=1}^{B}$ is valid by the monotonicity argument given below.

S.1.2 The Berk–Jones Weighting

Constrained thresholding removes coordinates but does not change how the surviving ones are compared. The plain maximum treats a deviation of a given size identically in a cell of 30 dyads and in a cell of 30,000, although under the null the first is routine and the second is extremely unlikely, so a common cutoff on raw deviations is effectively set by the smallest cells. The Berk–Jones weight replaces the raw deviation at each coordinate by a measure of how unlikely a deviation of that size is in a cell of that size, which places all cells on one scale.

Fix a candidate $\overline{\theta}$, a cell z_D with $\hat{m}(z_D)$ dyads, and a threshold c. Write $u := F_\varepsilon(c; \theta_\varepsilon)$. Under the null $H_0 : \overline{\theta}_0 = \overline{\theta}$, the number of dyads in the cell whose shock falls at or below c is

$$N(z_D, c) := \sum_{i<j:\, Z_{D,ij} = z_D} \mathbf{1}\{\varepsilon_{ij} \le c\} \ \Big|\ Z \ \sim\ \mathrm{Binomial}\big(\hat{m}(z_D),\, u\big)$$

by Assumptions 2 and 3, so the cell frequency $\hat{F}_n(c; z_D) = N(z_D, c)/\hat{m}(z_D)$ has conditional mean u. Write $m := \hat{m}(z_D)$ and let $\mathrm{KL}(p\|u) := p \log \frac{p}{u} + (1-p) \log \frac{1-p}{1-u}$ be the Bernoulli Kullback–Leibler divergence of the main text. For every m and every $p \in (u, 1)$ such that mp is an integer,

$$(m+1)^{-1} e^{-m\,\mathrm{KL}(p\|u)} \ \le\ \mathbb{P}\big(\hat{F}_n(c; z_D) \ge p \mid Z\big) \ \le\ e^{-m\,\mathrm{KL}(p\|u)}, \tag{S.1}$$

and symmetrically for $p \in (0, u)$ with the event $\{\hat{F}_n(c; z_D) \le p\}$. Both bounds are nonasymptotic, and no normal approximation enters at any point. The upper bound

follows from Markov's inequality applied to $e^{\lambda N}$ with $N \sim \text{Binomial}(m, u)$: for any $\lambda > 0$,

$$\mathbb{P}(N \geq mp) \;\leq\; e^{-\lambda mp}\big(1 - u + ue^{\lambda}\big)^{m} \;=\; \exp\Big\{ - m\big[\lambda p - \log(1 - u + ue^{\lambda})\big]\Big\},$$

and the bracket is maximized at $e^{\lambda} = \frac{p(1-u)}{u(1-p)}$, where it equals exactly $\mathrm{KL}(p\|u)$. The lower bound is the method-of-types estimate. With $H(p) := -p\log p - (1-p)\log(1-p)$, the binomial coefficient satisfies $\binom{m}{mp} \geq (m+1)^{-1}e^{mH(p)}$, and $u^{mp}(1-u)^{m(1-p)} = e^{-m[H(p)+\mathrm{KL}(p\|u)]}$, so $\mathbb{P}(N \geq mp) \geq \mathbb{P}(N = mp) \geq (m+1)^{-1}e^{-m\,\mathrm{KL}(p\|u)}$. Taking logarithms in (S.1) gives

$$m\,\mathrm{KL}(p\|u) \;\leq\; -\log \mathbb{P}\big(\hat{F}_n(c; z_D) \geq p \mid Z\big) \;\leq\; m\,\mathrm{KL}(p\|u) + \log(m+1),$$

so $m\,\mathrm{KL}(p\|u)$ is the exact exponent of the tail probability up to an additive logarithmic term. In the language of large deviations, $\log \mathbb{P} = -m\,\mathrm{KL}(p\|u) + O(\log m)$, but the bounds above are sharper than that asymptotic statement and hold at every finite m.

This is what it means to report an inequality on a tail-probability scale. Suppose that at (z_D, c) the lower envelope violates the sandwich, $\hat{p}_L^{(n)}(z_D, c; \theta) > F_\varepsilon(c)$, and consider the weighted score $\hat{m}(z_D)\,\mathrm{KL}\big(\hat{p}_L^{(n)} \,\big\|\, F_\varepsilon(c)\big)$. Since $\hat{m}(z_D)\hat{p}_L^{(n)}$ is an integer[1], (S.1) shows that the score equals, up to at most $\log(\hat{m}(z_D)+1)$, the negative logarithm of the conditional probability that the cell frequency $\hat{F}_n(c; z_D)$ deviates from $F_\varepsilon(c)$ at least as far as the observed envelope does. The score is thus minus the logarithm of a one-sided p-value for the pointwise null that the cell's frequency has mean $F_\varepsilon(c)$. A score of w in a cell of 30 dyads and a score of w in a cell of 30,000 dyads therefore describe deviations that are equally surprising under the null, even though the raw deviations differ enormously. The maximum over cells and thresholds then compares like with like and discounts the large raw deviations that small cells produce by chance. Equivalently, $\hat{m}(z_D)\,\mathrm{KL}\big(p \,\big\|\, F_\varepsilon(c)\big)$ is the binomial log-likelihood-ratio statistic for that pointwise null, so the maximum over c is the supremum-of-likelihood-ratio statistic of Berk and Jones (1979), applied cell by cell within the constrained windows.

At boundary frequencies we adopt the usual extended conventions $0\log 0 := 0$, $\mathrm{KL}(p\|0) := \infty$ for $p > 0$, $\mathrm{KL}(p\|1) := \infty$ for $p < 1$, and $\mathrm{KL}(0\|0) = \mathrm{KL}(1\|1) := 0$, under which the weighted scores remain well defined and the monotonicity verified

[1]Note that the argument remains valid without requiring $\hat{m}(z_D)\hat{p}_L^{(n)}$ to be an integer, by applying a Chernoff bound, albeit at the cost of some additional conservativeness.

next is preserved. The boundary cases of (S.1) follow by taking limits under the same conventions.

S.1.3 Validity of the Combined Aggregation

Both improvements are justified by Theorem 4 of the main paper, whose only requirement on the aggregation T is that it be nondecreasing in each coordinate of V, the coordinates being the nonnegative violations $V^{\pm}$. We verify this requirement for the windowed and weighted statistic $\widehat{Q}_n^{\mathrm{BJ}}$ of the main text. For $0 < F_\varepsilon(c) < 1$, the map $p \mapsto \mathrm{KL}\big(p \,\big\|\, F_\varepsilon(c)\big)$ is strictly decreasing on $[0, F_\varepsilon(c)]$ and strictly increasing on $[F_\varepsilon(c), 1]$, and under the boundary conventions above it remains nondecreasing away from $F_\varepsilon(c)$ when $F_\varepsilon(c) \in \{0, 1\}$, which is all that is needed. When $\hat{p}_L^{(n)} > F_\varepsilon(c)$ we may write $\hat{p}_L^{(n)} = F_\varepsilon(c) + V^+$, so the score at that coordinate is $\hat{m}(z_D)\,\mathrm{KL}\big(F_\varepsilon(c) + V^+ \,\big\|\, F_\varepsilon(c)\big)$, which is nondecreasing in V^+ and vanishes at $V^+ = 0$. Symmetrically, the score when $\hat{p}_U^{(n)} < F_\varepsilon(c)$ is $\hat{m}(z_D)\,\mathrm{KL}\big(F_\varepsilon(c) - V^- \,\big\|\, F_\varepsilon(c)\big)$, which is again nondecreasing in V^-. Each coordinate is thus passed through a nondecreasing function of its own violation, and constrained thresholding multiplies each coordinate by a (θ, Z)-measurable indicator, which is again nondecreasing. A maximum of nondecreasing functions is nondecreasing, so the aggregation T^{BJ} satisfies the requirement, Theorem 4 applies, and the CS obtained by inverting the test against the $\lceil (1-\alpha)(B+1) \rceil$-th smallest value among simulated copies of $R_n^{\mathrm{BJ}}(\overline{\theta})$ is valid. The same argument, with the identity in place of the Kullback–Leibler weight, covers $\widehat{Q}_n^{\mathrm{con}}$.

Two structural features of the weighted statistic are worth recording. First, the conditions selecting the first two branches of the discrepancy d_{BJ} are mutually exclusive at any (z_D, c), since $\hat{p}_L^{(n)} \le \hat{p}_U^{(n)}$, which holds at every θ because $(1 - Y_{ij})\mathbf{1}\{\delta_{ij}(\theta) \ge c\} + Y_{ij}\mathbf{1}\{\delta_{ij}(\theta) \le c\} \le 1$ for every dyad. Second, the same holds on the control side, where $\hat{F}_n(c; z_D)$ lies on one side of $F_\varepsilon(c)$ and the two-sided deviation therefore collapses to the single score $\hat{m}(z_D)\,\mathrm{KL}\big(\hat{F}_n(c; z_D) \,\big\|\, F_\varepsilon(c)\big)$. Taking the maximum over cells and the supremum over the constrained threshold sets gives $R_n^{\mathrm{BJ}}(\overline{\theta})$ displayed in the main text. Its simulated copies are obtained by replacing $\hat{F}_n$ with $\hat{F}_n^{(b)}$ and equal $T_{\overline{\theta}}^{\mathrm{BJ}}(R^{(b)})$.

S.1.4 Improved Semiparametric CS

When F_ε is unknown, the domination argument itself carries over to the semiparametric procedure of Section 3.1 of the main paper once the inequalities are indexed by ordered *pairs* of cells rather than by a cell and a sign. Formally, for $z_D, z'_D \in \hat{\mathcal{Z}}_D$ and $c \in \mathbb{R}$, set

$$V(z_D, z'_D, c; \theta) := \left[\hat{p}_L^{(n)}(z_D, c; \theta) - \hat{p}_U^{(n)}(z'_D, c; \theta)\right]_+,$$
$$R(z_D, z'_D, c) := \left[\hat{F}_n(c; z_D) - \hat{F}_n(c; z'_D)\right]_+.$$

At θ_0 the sandwich (9) gives $V \le R$ coordinate by coordinate, and the statistics (11) and (13) of the main paper are the special case $T = \max$. The argument of Theorem 4, combined with the transform step in the proof of Theorem 2 of the main paper, then yields a valid CS for any measurable nondecreasing $T_{Z,\theta}$ of the *supremum type*, meaning a supremum over c of a per-threshold aggregation that does not otherwise depend on c. Under the transform $u = F_\varepsilon(c)$, every value that such a T attains as c sweeps $\mathbb{R}$ is also attained by its uniform-scale counterpart at some $u \in F_\varepsilon(\mathbb{R})$, so the value at the truth is dominated by the supremum of that counterpart over $u \in [0,1]$, with equality when F_ε is continuous and with the distributional transform of the main text when it has atoms. The critical value is therefore simulated from $R^{(b)}(z_D, z'_D, u) := \left[G^{(b)}_{z_D}(u) - G^{(b)}_{z'_D}(u)\right]_+$ exactly as in Section 3.1.

The restriction to the supremum type is essential, and it marks the key difference from the parametric case: neither the threshold windows nor the Berk–Jones weights, both of which vary with c, carry over. The semiparametric test exploits the probability integral transform $\hat{F}_n(c; z_D) = G_{z_D}\big(F_\varepsilon(c)\big)$, with uniform simulated draws and the supremum over $c \in \mathbb{R}$ replaced by $\sup_{u\in[0,1]}$, which no longer involves c. A weight that depends on c would interfere with that transform and invalidate the simulation-based calibration. A window has a different issue. Its image $F_\varepsilon\big(C(z_D; \theta, Z)\big)$ on the uniform scale is unknown, so the simulated supremum must still run over all of $[0,1]$, and a windowed observable statistic compared with that unwindowed cutoff is valid but gains nothing over the unwindowed statistic. Weights may still depend on Z, in particular on the effective sample sizes $\hat{m}(z_D)$ and $\hat{m}(z'_D)$, so that restrictions coming from pairs of very small cells can still be downweighted.

S.1.5 Implementation in the Simulations

The semiparametric columns of Table 3 of the main paper and of Tables S.2 and S.3 below use one such weighting, which we now record. For each retained cell let

$$e(z_D) := \left[\frac{\log(2\hat{\kappa}/\alpha)}{2\hat{m}(z_D)}\right]^{1/2}, \qquad \hat{\kappa} := \#\hat{\mathcal{Z}}_D,$$

the radius at which the Dvoretzky–Kiefer–Wolfowitz inequality of Section S.4 assigns tail probability $\alpha/\hat{\kappa}$ to cell z_D; the constant is common to all cells, so $e(z_D) \propto \hat{m}(z_D)^{-1/2}$. The weighted criterion divides each cross-cell gap by the sum of the two radii,

$$\widehat{Q}_n^{\mathrm{w}}(\theta) := \sup_{c\in\mathbb{R}} \left[\max_{z_D, z_D' \in \hat{\mathcal{Z}}_D} \frac{\hat{p}_L^{(n)}(z_D, c; \theta) - \hat{p}_U^{(n)}(z_D', c; \theta)}{e(z_D) + e(z_D')} \right]_+,$$

and its simulated control statistic is the same functional of the uniform empirical CDFs, $R^{\mathrm{w},(b)} := \sup_{u\in[0,1]} \max_{z_D, z_D'} \left[G_{z_D}^{(b)}(u) - G_{z_D'}^{(b)}(u)\right] / \left[e(z_D) + e(z_D')\right]$. In the computations the supremum over u is replaced by a maximum over the grid $u_k = k/4{,}096$ at which the two cells of every pair are evaluated at adjacent grid points, $G_{z_D}^{(b)}(u_k) - G_{z_D'}^{(b)}(u_{k-1})$, so that the maximum over all pairs of cells weakly exceeds the exact supremum over the bin $(u_{k-1}, u_k]$ and is therefore conservative. This is a nondecreasing aggregation of the supremum type whose weights depend on Z only through the cell sizes, so the argument above applies: with $\hat{q}_{n,1-\alpha}^{\mathrm{w}}(Z)$ the $\lceil (1-\alpha)(B+1) \rceil$-th smallest of $R^{\mathrm{w},(1)}, \ldots, R^{\mathrm{w},(B)}$, the CS $\{\theta : \widehat{Q}_n^{\mathrm{w}}(\theta) \le \hat{q}_{n,1-\alpha}^{\mathrm{w}}(Z)\}$ has conditional coverage at least $1-\alpha$. In the computations the critical value is carried inside the criterion, as the multiple $\hat{q}_{n,1-\alpha}^{\mathrm{w}}(Z)$ of each cell's radius subtracted from its envelope, and the comparison is then against zero; the two forms are identical. A union bound over the retained cells, as in the proof of Proposition S.1, shows that setting that multiple to one is valid without any simulation; the finite-B simulated multiple is random and need not lie below one, but it typically does, and the resulting gain over the analytic value is largest at small n, where the union bound is loosest. The symmetry-restricted criterion of Section S.2 is weighted in the same way, term by term in each of the three terms of (S.2), with the mirror evaluations at $1-u$ carrying the radii of their own cells; its validity follows from the domination displays of that section applied to the weighted terms, since the weights depend only on the cell sizes.

S.2 Semiparametric Inference Based on Symmetry Restrictions

Section S.1 reweighted the semiparametric inequalities without adding any. This section instead adds inequalities: it adapts the semiparametric inference procedure of Section 3.1 of the main paper to *symmetry* restrictions on F_ε, which leave F_ε nonparametric and therefore place the resulting CS between the parametric and the fully semiparametric procedures of the main paper. We take F_ε to be continuous for simplicity.

Suppose F_ε is *symmetric about* 0, i.e.,

$$F_\varepsilon(-c) = 1 - F_\varepsilon(c), \quad \forall c \in \mathbb{R}.$$

Once symmetry is imposed, setting the center to be 0 is without loss of generality as a location normalization.[2] Given this observation, any parametric assumption that restricts F_ε to a known symmetric family of distributions, such as the normal, logistic, Laplace, and uniform, becomes a strict special case of the nonparametric symmetry assumption above.

Condition on Z. Write $\hat{a}(c) := \max_{z_D \in \hat{\mathcal{Z}}_D} \hat{p}_L^{(n)}(z_D, c; \theta)$ and $\hat{b}(c) := \max_{z_D \in \hat{\mathcal{Z}}_D} \left[1 - \hat{p}_U^{(n)}(z_D, c; \theta)\right]$. Define

$$\widehat{Q}_n^{\text{sym}}(\theta) := \sup_c \left[\max\left\{ \hat{a}(c) + \hat{b}(c),\ \hat{a}(c) + \hat{a}(-c),\ \hat{b}(c) + \hat{b}(-c) \right\} - 1 \right]_+. \tag{S.2}$$

The first term, corresponding to $\hat{a}(c) + \hat{b}(c) \leq 1$, encodes the semiparametric restriction already used in Section 3.1 of the main paper, while the other two are the mirror restrictions $\hat{a}(c) + \hat{a}(-c) \leq 1$ and $\hat{b}(c) + \hat{b}(-c) \leq 1$ motivated by the symmetry restriction $F_\varepsilon(-c) = 1 - F_\varepsilon(c)$.

Then at θ_0 each of the three terms is dominated, realization by realization, by the corresponding functional of the per-cell uniform empirical CDFs G_{z_D} evaluated at $u = F_\varepsilon(c)$ *and at its mirror* $1 - u$. Writing $D_{z_D}(u) := G_{z_D}(u) - u$, we have:

$$\hat{a}(c) + \hat{b}(c) - 1 \leq \max_{z_D} G_{z_D}(u) - \min_{z_D} G_{z_D}(u)$$
$$\hat{a}(c) + \hat{a}(-c) - 1 \leq \max_{z_D} D_{z_D}(u) + \max_{z_D} D_{z_D}(1 - u)$$

[2] Given the symmetry-about-0 assumption, a constant term must be included in Z_{ij} so that θ includes an unrestricted intercept component.

$$\hat{b}(c) + \hat{b}(-c) - 1 \leq -\min_{z_D} D_{z_D}(u) - \min_{z_D} D_{z_D}(1-u).$$

All three follow from the sandwich (9) together with the identity $\hat{F}_n(c; z_D) = G_{z_D}(F_\varepsilon(c))$ established in the proof of Theorem 2 of the main paper, which holds exactly and therefore calls for no left limits; for the last two we also use $F_\varepsilon(-c) = 1-u$, after which the terms u and $1-u$ cancel against the -1 on the left. Note that the first right-hand side is exactly the cross-cell range in (13) evaluated at u, so the symmetry-restricted criterion nests the procedure of Section 3.1 and simply adds the two mirror terms.

The right-hand sides, along with their suprema over u, again have conditional distributions (given Z) that do not depend on F_ε and can therefore be simulated exactly as in Section 3.1 of the main paper: draw $\hat{m}(z_D)$ i.i.d. uniform random variables in each cell to form $G^{(b)}_{z_D}$, and take the maximum over $u \in [0,1]$. Concretely, writing $D^{(b)}_{z_D}(u) := G^{(b)}_{z_D}(u) - u$, the single joint simulated statistic is

$$R^{(b)}_{\text{sym}} := \sup_{u\in[0,1]} \max\Big\{ \max_{z_D} G^{(b)}_{z_D}(u) - \min_{z_D} G^{(b)}_{z_D}(u),\ \max_{z_D} D^{(b)}_{z_D}(u) + \max_{z_D} D^{(b)}_{z_D}(1-u), \\ - \min_{z_D} D^{(b)}_{z_D}(u) - \min_{z_D} D^{(b)}_{z_D}(1-u)\Big\}.$$

By the three displays above, $\widehat{Q}^{\text{sym}}_n(\theta_0)$ is dominated, realization by realization, by the same functional of the true per-cell uniform empirical CDFs, which shares the conditional distribution of $R^{(b)}_{\text{sym}}$. Comparing $\widehat{Q}^{\text{sym}}_n(\theta)$ with the finite-B-adjusted $(1-\alpha)$-quantile of the simulated draws and inverting the test then yields level-$(1-\alpha)$ conditional CSs. In the computations each of the three terms is weighted cell by cell as described at the end of Section S.1. Section S.3.3 reports the finite-sample performance of this procedure on the simulation design of the main paper.

S.3 Additional Simulation Results

This section collects three further numerical exercises based on the simulation design of Section 4.1 of the main paper: a shock-dependence exercise covering both theory-covered and robustness designs, a repetition of the main exercise with multidimensional exogenous covariates, and the performance of the symmetry-restricted semiparametric procedure of Section S.2. All three exercises use $K = 500$ replications at the 95% nominal level. Throughout, "unrestricted" denotes the fully semiparamet-

ric procedure of Section 3.1 of the main paper, which leaves F_ε unrestricted. Each table reports coverage, either as a column or, when the true γ grid point is covered in every replication, in the table notes, together with the informativeness diagnostics of the main text: the sign rate, the median width, and, where it is reported as a column, the vacuity share of the γ projection. Width is the total length of the accepted γ projection and every table reports its median over replications. A median width is truncated by the candidate grid whenever the accepted γ projection touches an endpoint of that grid, and we note below where this happens often enough to matter.

S.3.1 Common and Dyadic Shock Dependence

Remark 1 of the main paper states that an additive common shock is absorbed by the free intercept in the parametric procedure. We examine that implication, and then depart from the maintained assumptions, using the parametric CS at $n = 1{,}600$ under three error processes. Arm A is the i.i.d. standard-logistic baseline. Arm B adds a replication-level common shock $\eta \sim N(0, \sigma_\eta^2)$ with $\sigma_\eta = 0.15$ to otherwise i.i.d. logistic shocks, so that the coverage target is the slope vector together with the shock-shifted intercept $b_0 + \eta$. The scale keeps the realized densities, between 10.76% and 36.38% with a mean of 20.59%, comparable to the baseline's (18.26% to 23.57%), so that the arm isolates the effect of the shift itself. Arm C preserves standard-logistic marginals but generates the shocks through a Gaussian copula with a group-pair factor: the n nodes are split into $G = 4$ equal groups, and the shock of a dyad between groups a and b is the logistic quantile of the normal cdf of $\sqrt{\tau}\, g_{ab} + \sqrt{1-\tau}\, w_{ij}$, with i.i.d. standard normal g_{ab} and w_{ij} and factor share $\tau = 0.65$. Arms A and B are covered by the maintained theory, whereas Arm C is a robustness exercise that breaks dyadic independence without changing the marginal distribution of any single shock.

Table S.1 reports the results. Where the theory applies, the additive shock is absorbed at almost no cost: the γ projection and the full parameter vector are covered in every replication of Arms A and B, both arms sign γ in every replication, and the median width is 12.50 under the shock against 12.25 without it. Outside the guarantee the γ projection degrades visibly while the full parameter vector loses most of its coverage. Under block dependence the γ projection covers its true value in 86.0% of the replications, the b_0 and β projections in 73.2% and 84.0%, and the full parameter vector in 12.4%. Most γ-coverage misses are sets lying entirely above $\gamma_0 = 16$, as

Table S.1: Finite-sample performance under shock dependence: parametric confidence sets at $n = 1{,}600$

	coverage			
error process	γ	joint	sign rate	median width
i.i.d. baseline	1.000	1.000	1.000	12.25
additive common shock, $\sigma_\eta = 0.15$	1.000	1.000	1.000	12.50
block copula, $\tau = 0.65$, $G = 4$	0.860	0.124	0.984	31.50

Notes: $n = 1{,}600$, $K = 500$, $B = 999$, γ grid $[-16, 72]$ in steps of 0.25. "γ" coverage refers to the γ projection and "joint" coverage to the full parameter vector (b_0, β, γ); sign rate and width refer to the γ projection.

one expects when a dense group pair generates common friends whose effect the model attributes to γ, and five replications accept the empty set.[3] The high sign rate (98.4%) does not resolve the coverage failure: sets lying entirely above γ_0 exclude both zero and the true parameter. Under group-pair (block copula) dependence, network density ranges from 2.57% to 62.53% across replications. Coverage failures in this case concentrate in sparse networks, with coverage of γ_0 falling to 71.8% (51 of 71 replications) when density is below 10%. The median projection width rises to 31.50, and the projection reaches a boundary of the length-88 candidate grid in 22.6% of replications.

S.3.2 Multidimensional Exogenous Covariates

The simulation design of Section 4.1 of the main paper has one exogenous covariate $|z_i - z_j|$. We now repeat the exercise with two and three exogenous covariates, $Z_{ij} = (|z_{i1} - z_{j1}|, |z_{i2} - z_{j2}|)$ and $Z_{ij} = (|z_{i1} - z_{j1}|, |z_{i2} - z_{j2}|, |z_{i3} - z_{j3}|)$, with true coefficient vectors $\beta_0 = (-1, +1)$ and $\beta_0 = (-1, +1, -0.5)$, the same strategic coefficient $\gamma_0 = 16$ on $\overline{\mathrm{CF}}_{ij}$, and standard-logistic shocks. Each margin of z_i is independently uniform on the same 21 equally spaced points in $[-10, 10]$, so the exogenous index is spread over a range two to two and a half times wider than in the scalar design. The intercept is calibrated once on pilot seeds, disjoint from the production seeds, so that the networks at $n = 400$ have a mean density of 20%, and is then frozen. Because the exogenous index depends on (z_i, z_j) only through the distance vector Z_{ij}, dyads are

[3] The median width of the block-copula arm is taken over the 495 nonempty sets.

grouped by that vector, which takes $21^2 = 441$ and $21^3 = 9{,}261$ values. Pooling node-type pairs that share a distance vector is a valid, deliberately coarser partition in the sense of Definition 1 of the main paper. The unrestricted semiparametric procedure conditions on these exact distance-vector cells with a floor of $\underline{m} = 20$ dyads. The parametric procedure, whose cost grows with the number of cells times the number of candidates, conditions on a coarser partition that groups the 21 distance values of each margin into seven consecutive bins, giving at most 49 and 343 cells, and retains every nonempty cell. We use coarser candidate grids than in the main text to limit computational cost. The γ grid is $[-16, 64]$ in steps of 1 for both procedures, and the parametric procedure evaluates the intercept and β on a lattice with a handful of points per coordinate that contains the truth. The semiparametric procedure fixes $b_0 = 0$ as a location normalization and $|\beta_1| = 1$ as a scale normalization, takes the union of the two sign charts, and sweeps the remaining slope coefficients in steps of 0.5. Both critical values are simulated with $B = 999$ as in the main text, the semiparametric one under the cellwise calibration of Section S.1. The results are reported in Table S.2.

Three features of the table stand out. First, coverage is exact: the true γ grid point is included in every replication of every entry. Second, sign revelation is fast and complete. The parametric procedure certifies the positive sign of γ_0 in 73.0% and 46.0% of the replications at $n = 100$ in the two- and three-dimensional designs, in more than 90% at $n = 200$, and in every replication from $n = 400$ onward, exactly as in the scalar design. The semiparametric procedure certifies the sign in every replication from $n = 1{,}600$ onward in both designs, having done so in 23.4% and 4.6% of the replications at $n = 800$, so the two procedures are again ordered by the strength of the maintained assumptions. Third, the parametric γ projection contracts quickly at small n and then levels off, at a median width of 24.00 on the grid of length 80 in the two-dimensional design and 20.00 in the three-dimensional design from $n = 6{,}400$ onward, whereas the scalar design of the main text reaches 2.25 on its grid of length 88 at $n = 10{,}000$. The broad projections persist under the parameter-grid refinements examined. Coarse conditioning on Z is a plausible contributor: pooling distinct distance vectors weakens the restrictions, and this information loss persists as sample size increases. With greater computational resources, finer parameter searches and conditioning partitions should improve numerical accuracy and may yield more informative confidence sets. We therefore read this table primarily for its sign content,

Table S.2: Multidimensional exogenous covariates: γ projections

dim-Z	n	parametric			semiparametric		
		sign	width	vac	sign	width	vac
2	100	0.730	34.00	0.000	0.000	80.00	1.000
	200	0.992	30.00	0.000	0.000	80.00	0.962
	400	1.000	27.00	0.000	0.002	78.00	0.254
	800	1.000	26.00	0.000	0.234	56.00	0.000
	1,600	1.000	25.00	0.000	1.000	35.00	0.000
	3,200	1.000	25.00	0.000	1.000	29.00	0.000
	6,400	1.000	24.00	0.000	1.000	25.00	0.000
	10,000	1.000	24.00	0.000	1.000	25.00	0.000
3	100	0.460	39.00	0.002	0.000	80.00	1.000
	200	0.916	31.00	0.000	0.000	80.00	1.000
	400	1.000	26.00	0.000	0.000	80.00	0.998
	800	1.000	23.00	0.000	0.046	77.50	0.284
	1,600	1.000	22.00	0.000	1.000	61.00	0.000
	3,200	1.000	21.00	0.000	1.000	40.50	0.000
	6,400	1.000	20.00	0.000	1.000	32.00	0.000
	10,000	1.000	20.00	0.000	1.000	29.00	0.000

Notes: Columns as in Table 3 of the main paper. The true γ_0 is covered in every replication.

which is unambiguous.

At small n the semiparametric γ projections are mostly vacuous, because the distance-vector cells above the floor of 20 dyads are thin and their allowances correspondingly wide. In the three-dimensional design no cell reaches the floor at $n = 100$, so the CS is the full grid by convention, and the projections remain vacuous in every replication at $n = 200$, in 99.8% at $n = 400$, and in 28.4% of them at $n = 800$. The accepted projection touches a grid endpoint in more than 5% of the replications through $n = 800$ in the two-dimensional design and through $n = 1{,}600$ in the three-dimensional design. By $n = 10{,}000$ the semiparametric median width has contracted to 25.00 in the two-dimensional design and 29.00 in the three-dimensional design.

S.3.3 Scalar Semiparametric Inference under Symmetry

We evaluate the symmetry-restricted semiparametric procedure on the simulation design of the main paper, on the same simulated networks and the same γ grid on $[-16, 72]$ in steps of 0.125. The parametric column uses the conditional simulation critical values of the main text, and both semiparametric columns use the cellwise calibration of Section S.1, so that all three columns are calibrated by conditional simulation at $B = 999$. The symmetry-restricted procedure fixes the scale normalization $|\beta| = 1$, takes the union of the two sign charts of the normalized coefficient, and profiles the unknown center of symmetry, which is equivalent to sweeping a free intercept jointly with γ. Table S.3 reports the sign rate, median width, and vacuity share of the γ projections, alongside the parametric and the fully semiparametric (semiparametric) CSs of the main text.

Two patterns stand out. First, the symmetry-restricted procedure certifies the sign of γ_0 in every replication from $n = 800$ onward. By comparison, the fully semiparametric procedure certifies the sign in 99.8% (499 of 500) of replications from $n = 6{,}400$ onward, while the parametric procedure achieves full sign certification from $n = 400$ onward. Second, the parametric CSs have the smallest median width at every network size. Symmetry does not uniformly reduce widths relative to the fully semiparametric procedure. The symmetry-restricted median widths are larger from $n = 200$ through $n = 3{,}200$ and smaller from $n = 6{,}400$ onward. At $n = 10{,}000$, the median widths are 2.25, 12.00, and 12.50 for the parametric, symmetry-restricted, and fully semiparametric CSs, respectively. This uneven width ordering may reflect criterion-specific calibration, which can offset the tightening from additional restrictions. The clearest benefit of symmetry in this design is therefore earlier sign revelation, with the true γ_0 covered in every replication.

S.4 Analytic Critical Values

As an alternative to the simulated critical value of the main paper, an analytic critical value can be derived from a bound on the tail probability of R_n:

Proposition S.1 (Analytic Critical Value)**.** *Suppose Assumptions 1–3 of the main*

Table S.3: Semiparametric confidence sets under the symmetry: γ projections

n	parametric			symmetric			semiparametric		
	sign	width	vac	sign	width	vac	sign	width	vac
100	0.748	48.31	0.074	0.570	71.38	0.096	0.416	72.00	0.094
200	0.948	33.44	0.012	0.858	68.31	0.012	0.560	67.63	0.018
400	1.000	24.88	0.000	0.982	57.19	0.000	0.666	53.94	0.000
800	1.000	18.13	0.000	1.000	42.00	0.000	0.716	40.00	0.000
1,600	1.000	12.63	0.000	1.000	24.88	0.000	0.868	23.38	0.000
3,200	1.000	7.75	0.000	1.000	21.25	0.000	0.978	20.00	0.000
6,400	1.000	2.88	0.000	1.000	15.38	0.000	0.998	16.25	0.000
10,000	1.000	2.25	0.000	1.000	12.00	0.000	0.998	12.50	0.000

Notes: The parametric and semiparametric columns are those of Table 3 of the main paper. The true γ_0 is covered in every replication.

paper hold. Then for any $t > 0$,

$$\mathbb{P}\big(R_n > t \mid Z\big) \;\le\; \sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-\hat{m}(z_D)t^2/2}.$$

Consequently, let $\hat{t}_n(\alpha; Z)$ be the solution[4] *to*

$$\sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-\hat{m}(z_D)\hat{t}_n^2/2} = \alpha. \tag{S.3}$$

Then the CS $\{\theta : \widehat{Q}_n(\theta) \le \hat{t}_n(\alpha; Z)\}$ has conditional coverage at least $1 - \alpha$.[5]

Proof of Proposition S.1. Condition on Z and take any two cells $z_D, z_D' \in \hat{\mathcal{Z}}_D$. For any c,

$$\hat{F}_n(c; z_D) - \hat{F}_n(c; z_D') = \big[\hat{F}_n(c; z_D) - F_\varepsilon(c)\big] - \big[\hat{F}_n(c; z_D') - F_\varepsilon(c)\big] \;\le\; 2\tilde{R}_n,$$

[4] With $\hat{\kappa} = \#\hat{\mathcal{Z}}_D$ as in Section S.1, when $\hat{\kappa} \ge 1$ the left-hand side of (S.3) decreases continuously and strictly from $2\hat{\kappa}$ to 0 as $\hat{t}_n$ runs over $[0, \infty)$, so a solution exists and is unique for every $\alpha \in (0, 1)$. On the event $\hat{\mathcal{Z}}_D = \varnothing$ we set $\hat{t}_n(\alpha; Z) := 0$; by the convention of the main text all statistics are then zero and the CS is the full candidate parameter space.

[5] One can also work with the more conservative closed-form cutoff $\hat{t}_n = \sqrt{2\log(2\hat{\kappa}/\alpha) \,/\, \min_{z_D \in \hat{\mathcal{Z}}_D} \hat{m}(z_D)}$, obtained by replacing every $\hat{m}(z_D)$ by the smallest one, which does not require solving (S.3).

where

$$\tilde{R}_n := \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c \in \mathbb{R}} \big|\hat{F}_n(c; z_D) - F_\varepsilon(c)\big|.$$

It follows that $R_n \le 2\tilde{R}_n$, and hence

$$\mathbb{P}(R_n > t \mid Z) \le \mathbb{P}(\tilde{R}_n > t/2 \mid Z).$$

Conditional on Z, $\hat{F}_n(\cdot\,; z_D)$ is the empirical distribution function of $\hat{m}(z_D)$ i.i.d. draws from F_ε in cell z_D. The Dvoretzky–Kiefer–Wolfowitz (DKW) inequality (Dvoretzky et al., 1956; Massart, 1990), which holds for every distribution function,[6] gives

$$\mathbb{P}\big(\sup_c |\hat{F}_n(c; z_D) - F_\varepsilon(c)| > s \mid Z\big) \le 2e^{-2\hat{m}(z_D)s^2}$$

for each cell z_D. Using the union bound over all cells $z_D \in \hat{\mathcal{Z}}_D$ yields

$$\mathbb{P}(\tilde{R}_n > s \mid Z) \le \sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-2\hat{m}(z_D)s^2}.$$

Setting $s := t/2$, we then have

$$\mathbb{P}\big(R_n > t \mid Z\big) \;\le\; \mathbb{P}(\tilde{R}_n > t/2 \mid Z) \;\le\; \sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-2\hat{m}(z_D)(t/2)^2} \;=\; \sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-\hat{m}(z_D)t^2/2}.$$

By the definition of $\hat{t}_n(\alpha; Z)$,

$$\mathbb{P}\big(R_n > \hat{t}_n(\alpha; Z) \mid Z\big) \;\le\; \sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-\hat{m}(z_D)\hat{t}_n^2(\alpha;Z)/2} = \alpha.$$

Combining the above with $\widehat{Q}_n(\theta_0) \le R_n$ from Lemma 2 of the main paper again yields the coverage guarantee. □

The proof uses only $\widehat{Q}_n(\theta_0) \le 2\tilde{R}_n$, so the cutoff $\hat{t}_n(\alpha; Z)$ is valid for any criterion that is dominated, realization by realization, by twice the maximal cellwise Kolmogorov–Smirnov deviation. In particular it is valid for the symmetry-restricted criterion $\widehat{Q}_n^{\mathrm{sym}}$ of Section S.2, each of whose three terms is so dominated by the displays there. For the parametric criterion the factor two is not needed, since Lemma 3

[6] No continuity of F_ε is needed. Writing U_{ij} for the distributional transforms of the shocks and G_{z_D} for their empirical CDF in cell z_D, the proof of Theorem 2 of the main paper gives $\hat{F}_n(c; z_D) = G_{z_D}(F_\varepsilon(c))$ exactly, so $\sup_c |\hat{F}_n(c; z_D) - F_\varepsilon(c)| \le \sup_{u \in [0,1]} |G_{z_D}(u) - u|$, which reduces the bound to the uniform case. The two-sided form with constant 2 holds for every $s > 0$, trivially so when $2e^{-2\hat{m}(z_D)s^2} \ge 1$.

of the main paper gives $\widehat{Q}_n^{\text{par}}(\overline{\theta}_0) \leq \tilde{R}_n$ directly, so the root of $\sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-2\hat{m}(z_D)t^2} = \alpha$, which equals $\hat{t}_n(\alpha; Z)/2$, is a valid analytic cutoff there.

Remark S.1 (Convergence Rate of Critical Values). Proposition S.1 not only provides us with a valid inference procedure, but also offers a theoretical result about the rate at which the critical value shrinks, since (S.3) pins down the rate at which $\hat{t}_n(\alpha; Z) \to 0$ as $n \to \infty$. To see this, write $m_{\min} := \min_{z_D \in \hat{\mathcal{Z}}_D} \hat{m}(z_D)$ and recall $\hat{\kappa} = \#\hat{\mathcal{Z}}_D$. Keeping only the smallest cell, and then bounding every cell by it, (S.3) implies $2\log(2/\alpha)/m_{\min} \leq \hat{t}_n^2 \leq 2\log(2\hat{\kappa}/\alpha)/m_{\min}$. Since $1 \leq \hat{\kappa} \leq \kappa$ always, both bounds are $m_{\min}^{-1}$ times a constant. Under Assumption 4 of the main paper, $\hat{m}(z_D)$ is a U-statistic with the bounded symmetric kernel $\mathbf{1}\{(Z_i, Z_j) \in \mathcal{Z}_{D,k}\}$, so the strong law of large numbers for U-statistics gives $\hat{m}(z_D)/\binom{n}{2} \to q(z_D) := \mathbb{P}\left((Z_i, Z_j) \in \mathcal{Z}_{D,k}\right) > 0$ almost surely for each of the κ cells of Definition 1. Hence, $\hat{m}(z_D) \geq \underline{m}$ for all large n, so $\hat{\kappa} = \kappa$ eventually and $m_{\min} = \Theta(n^2)$ almost surely, and the critical value $\hat{t}_n(\alpha; Z)$ shrinks at the *dyadic* rate $\Theta(1/n)$, which is much faster than the rate $1/\sqrt{n}$ associated with the number of agents n. This is intuitive since our critical value is entirely built upon the "middle term" that involves ε_{ij} only (conditional on Z), and thus the $O(n^2)$ number of i.i.d. dyadic shocks ε_{ij} naturally induces a convergence rate of $\sqrt{1/n^2} = 1/n$ after averaging.

References

Athey, S., D. Eckles, and G. W. Imbens (2018): "Exact p-values for network interference," *Journal of the American Statistical Association*, 113, 230–240.

Auerbach, E. (2022a): "Identification and estimation of a partially linear regression model using network data," *Econometrica*, 90, 347–365.

——— (2022b): "Testing for differences in stochastic network structure," *Econometrica*, 90, 1205–1223.

Badev, A. (2021): "Nash Equilibria on (Un)Stable Networks," *Econometrica*, 89, 1179–1206.

Barnard, G. A. (1963): "Discussion of Professor Bartlett's paper," *Journal of the Royal Statistical Society: Series B (Methodological)*, 25, 294.

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

Berk, R. H. and D. H. Jones (1979): "Goodness-of-Fit Test Statistics That Dominate the Kolmogorov Statistics," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47, 47–59.

Canay, I. A., J. P. Romano, and A. M. Shaikh (2017): "Randomization tests under an approximate symmetry assumption," *Econometrica*, 85, 1013–1030.

Chesher, A. (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78, 575–601.

Chesher, A. and A. M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85, 959–989.

Comola, M. and A. Dekel (2026): "Estimating Network Externalities in Undirected Link Formation Games," *Journal of Applied Econometrics*, forthcoming.

de Paula, Á. (2020): "Econometric Models of Network Formation," *Annual Review of Economics*, 12, 775–799.

de Paula, Á., S. Richards-Shubik, and E. Tamer (2018): "Identifying Preferences in Networks With Bounded Degree," *Econometrica*, 86, 263–288.

de Paula, Á. and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games With Incomplete Information," *Econometrica*, 80, 143–172.

Dufour, J.-M. (2006): "Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics," *Journal of Econometrics*, 133, 443–477.

Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1956): "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator," *Annals of Mathematical Statistics*, 27, 642–669.

Dwass, M. (1957): "Modified randomization tests for nonparametric hypotheses," *Annals of Mathematical Statistics*, 28, 181–187.

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

Eagleson, G. K. and N. C. Weber (1978): "Limit Theorems for Weakly Exchangeable Arrays," *Mathematical Proceedings of the Cambridge Philosophical Society*, 84, 123–130.

Galichon, A. and M. Henry (2011): "Set Identification in Models With Multiple Equilibria," *The Review of Economic Studies*, 78, 1264–1298.

Gao, W. Y. (2020): "Nonparametric Identification in Index Models of Link Formation," *Journal of Econometrics*, 215, 399–413.

Gao, W. Y., M. Li, and S. Xu (2023): "Logical Differencing in Dyadic Network Formation Models With Nontransferable Utilities," *Journal of Econometrics*, 235, 302–324.

Gao, W. Y., M. Li, and Z. Xu (2026): "Tractable Identification of Strategic Network Formation Models With Unobserved Heterogeneity," arXiv:2603.08634.

Gao, W. Y. and R. Wang (2026): "Identification in Nonlinear Dynamic Panel Models under Partial Stationarity," *Journal of Econometrics*, 253, 106185.

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

Graham, B. S. and A. Pelican (2020): "Testing for Externalities in Network Formation Using Simulation," in *The Econometric Analysis of Network Data*, ed. by B. S. Graham and Á. de Paula, Cambridge, MA: Academic Press, 63–82.

Gualdani, C. (2021): "An Econometric Model of Network Formation With an Application to Board Interlocks Between Firms," *Journal of Econometrics*, 224, 345–370.

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

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

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

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

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

——— (2022): "Causal Inference Under Approximate Neighborhood Interference," *Econometrica*, 90, 267–293.

——— (2023): "Network Cluster-Robust Inference," *Econometrica*, 91, 641–667.

Leung, M. P. and H. R. Moon (2026): "Normal Approximation in Large Network Models," *The Review of Economic Studies*, forthcoming.

Li, L. and M. Henry (2026): "Finite Sample Inference in Incomplete Models," arXiv:2204.00473.

Li, M., Z. Shi, and Y. Zheng (2026): "Bagging the Network," arXiv:2410.23852.

Liu, W., Á. de Paula, and E. Tamer (2026): "Prediction Sets and Conformal Inference With Interval Outcomes," arXiv:2501.10117.

Manski, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

——— (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

Massart, P. (1990): "The Tight Constant in the Dvoretzky–Kiefer–Wolfowitz Inequality," *Annals of Probability*, 18, 1269–1283.

Mastrandrea, R., J. Fournet, and A. Barrat (2015): "Contact Patterns in a High School: A Comparison Between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys," *PLOS ONE*, 10, e0136497.

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

Menzel, K. (2021): "Central Limit Theory for Models of Strategic Network Formation," Working paper, New York University.

——— (2026): "Strategic Network Formation With Many Agents," *Journal of Econometrics*, 253, 106174.

Miyauchi, Y. (2016): "Structural Estimation of Pairwise Stable Networks With Nonnegative Externality," *Journal of Econometrics*, 195, 224–235.

Pelican, A. and B. S. Graham (2022): "An Optimal Test for Strategic Interaction in Social and Economic Network Formation Between Heterogeneous Agents," NBER Working Paper 27793; arXiv:2009.00212.

Ridder, G. and S. Sheng (2025): "Two-Step Estimation of a Strategic Network Formation Model With Clustering," arXiv:2001.03838.

Rosen, A. M. and T. Ura (2025): "Finite Sample Inference for the Maximum Score Estimand," *The Review of Economic Studies*, 92, 4117–4151.

Rozemberczki, B., C. Allen, and R. Sarkar (2021): "Multi-Scale Attributed Node Embedding," *Journal of Complex Networks*, 9, cnab014.

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

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

Wu, S. (2024): "Two-Step Estimation of Network Formation Models With Unobserved Heterogeneities and Strategic Interactions," arXiv:2404.12581.

Zeleneev, A. (2026): "Identification and Estimation of Network Models With Nonparametric Unobserved Heterogeneity," arXiv:2602.06885.