arXiv:2607.27505v1 [econ.EM] 29 Jul 2026

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

Wayne Yuan Gao† Ming Li‡

July 31, 2026

Abstract

We develop a finite-sample valid inference procedure for strategic network formation models in which linking decisions depend on endogenous network statistics (say, the number of common friends). Only a single network is required to be observed, and we restrict neither its density, nor the dependence structure induced by strategic interaction, nor the equilibrium selection mechanism. We exploit a *bounding-by-c* technique to construct a set of sandwich inequalities that are valid realization by realization, with the middle term involving only the i.i.d. pairwise error. We then average the sandwich inequalities over cells of exogenous covariates, and obtain identifying restrictions under a nonstandard *pathwise limit* formulation. For inference, we construct test statistics whose finite-sample uncertainty can be controlled by statistics of the exogenous covariates and errors alone, whose conditional distributions are exactly simulable in both semiparametric and parametric settings. Our proposed inference procedure is also computationally tractable, with *no need* to solve, simulate, or enumerate equilibrium network structures. In simulations, our procedure easily scales to networks of size $10,000$, and yields confidence sets that certifies the sign of the strategic coefficient. In two empirical applications (with network size about 300∼9500), we find statistical evidence for positive link interdependence at 95% confidence level.

Keywords: finite-sample inference, network formation, strategic interaction, single network, network dependence, partial identification, equilibrium multiplicity

*We thank Giovanni Compiani, Jiaying Gu, and Lixiong Li for helpful comments and suggestions.

†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 to well-positioned partners, for various reasons related to information, enforcement, and coordination. Estimating structural models that take this interdependence seriously is an important problem in the econometrics of network formation (de Paula, 2020).

Taking that interdependence seriously, however, creates at least three layers of challenges that compound each other. The first is the *equilibrium* itself. We adopt pairwise stability (Jackson and Wolinsky, 1996) as the solution concept, which is the leading equilibrium notion in the economics and econometrics literature on strategic network formation. The mapping from primitives to realized pairwise stable networks is intractable to characterize: the outcome space is combinatorially vast,[1] pairwise stability generically admits many equilibria at once (de Paula, 2020), and iterative procedures are not generally guaranteed to reach one even when it exists (Jackson and Watts, 2002). There is thus no tractable reduced form to invert. The second is the *dependence* this induces: each dyad's endogenous statistic is a function of the realized network, so every link can implicitly depend on every shock, rendering the dependence structure global and vastly complicated. On a *single* large network, the network dependence issue renders standard statistics tool such as Law of Large Numbers, Central Limit Theorem, bootstrap and subsampling highly nontrivial, whose validity often relies on complicated and hard-to-verify weak dependence conditions that equilibrium interdependence may destroy. The third is *computation*: procedures built on stability conditions must simulate shocks, search over completions of the observed network, and enumerate subnetwork configurations up to isomorphism. The endogenous covariate aggravates all three at once, since it must be recomputed at every candidate parameter and every simulated network, placing an equilibrium solver *inside* the inference loop rather than beside it.

This paper, to the best of our knowledge, is the first in the network 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 canonical one, where a link between

[1] For a standard illustration: there are 2^{435} possible undirected unweighted networks on 30 agents.

agents i and j forms if and only if the latent surplus exceeds a threshold:

$$Y_{ij} = \mathbf{1}\left\{Z'_{ij}\beta_0 + X'_{ij}\gamma_0 \geq \varepsilon_{ij}\right\}, \quad \text{for any } ij$$

where Z_{ij} collects exogenous observed dyadic covariates built from the two agents' characteristics, ε_{ij} is an idiosyncratic shock independent of them, and $X_{ij} := \phi_{ij}(Y, Z)$ collects *endogenous* observed network statistics (e.g. the number of common friends of i and j, the number of 2nd-order friends), which can be determined jointly by the equilibrium network Y (as well as the exogenous covariates Z). We assume that the observed network Y is pairwise stable under transferable utilities, or in other words, is a solution to the model equation above, while we leave entirely unrestricted the equilibrium selection mechanism. The model above corresponds to the transferable utility setting of Sheng (2020) and, with nonnegative externalities, of Miyauchi (2016). We note that our strategy is not confined to the transferable utility setting: Remark 2 explains how our key result extends to non-transferable utility settings. The structural parameter is $\theta_0 = (\beta_0, \gamma_0)$, with γ_0 being the key object of interest: it measures the strength of the strategic interdependence in link formation, and the null $\gamma_0 = 0$ is the hypothesis that there is none.[2]

The key device that enables us to handle the intractable network endogeneity is a technique we call *bounding by c*.[3] A formed link certifies that its idiosyncratic shock lies below its latent index, $\varepsilon_{ij} \leq Z'_{ij}\beta_0 + X'_{ij}\gamma_0$, and an absent link certifies the reverse. The certificate is not directly usable, because the index contains the endogenous X_{ij}. However, on the *observable* (random) event that the index does not exceed a deterministic scanning threshold, $Z'_{ij}\beta_0 + X'_{ij}\gamma_0 \leq c$, transitivity of the inequalities implies $\varepsilon_{ij} \leq c$, an event involving only the exogenous shock and the number c. Averaging such "bounding-by-c" certificates within a cell of exogenous characteristics therefore sandwiches the empirical distribution of that cell's exogenous shocks between two *observable* and *endogenous* envelopes, *at every n and in every realization*, before any expectation or limit is taken. Importantly, the *realization-wise* nature of the inequalities implies their validity at every possible equilibrium regardless of the equilibrium selection mechanism, with the conditional distribution of the endogenous covari-

[2] That null has itself been the object of a dedicated testing literature. Graham and Pelican (2020); Pelican and Graham (2022) construct exact tests of no strategic interaction by conditioning on the sufficient statistics the model admits under the null, where it collapses to an exponential-family dyadic model. Such tests are sharp against the no-interdependence null, but they do not invert into confidence sets for γ_0. Our confidence sets deliver the no-interdependence test as a by-product.

[3] This technique was proposed and adopted in Gao and Wang (2026), who studies sharp identification in dynamic binary choice models. That setting is very different from the strategic network formation model considered here. Gao and Wang (2026) incorporates individual-level fixed effects in a panel data setting without cross-sectional strategic interdependence, while our paper considers a one-shot strategic network formation game without fixed effects.

ates X_{ij} never modeled, estimated, or simulated. Identification and inference are then two ways of using the same realization-wise *bounding-by-c* inequalities obtained: taking limits along the realized network sequence gives identifying restrictions, while characterizing the finite-sample uncertainty of an empirical average of the exogenous shocks ϵ and exogenous covariates Z yields the valid non-asymptotic inference.

Specifically, for *identification*, we take finite-sample averages of the *bounding-by-c* inequalities, and consider the limit restrictions arising from a realized sequence of networks as $n \to \infty$. Since we do not impose any sparsity and dependence restrictions, joint probabilities of observable events that involve the endogenous covariates X *are not guaranteed to converge* in the single large-network limit. A key novelty of our identification result lies in that the validity of our identifying restriction only rests on the large-n convergence of an exogenous statistic involving ε and Z only, and thus does not require convergence of probabilities that involve the endogenous X. Consequently, our large-network identifying restrictions are stated in terms of *pathwise limit frequencies*: the limit superior of the lower envelope and the limit inferior of the upper envelope, each of which involves the endogenous X. This departs from the standard identification template based on (conditional) expectations or probabilities, in a way we think is of independent theoretical interest. No observable quantity is required to settle down to a deterministic population limit, and along a single network sequence the observable frequencies may fluctuate indefinitely and their limit envelopes may remain genuinely random, capturing lack of determination from equilibrium selection or other forms of strong global dependence.

For *finite-sample inference*, the paper's central contribution, we exploit use the same sandwich inequalities with n held fixed. Again, due to the realization-wise nature of our *bounding-by-c* inequalities, at the true θ_0, the observable criterion is bounded, realization by realization, by a dominance statistic defined on the exogenous shocks ε and the exogenous covariates Z alone. Importantly, this statistic involves no endogenous covariate X at all, and thus its distribution is not contaminated by the interdependence generated by the equilibrium network. We provide both semiparametric and parametric inference procedures: our key idea, in either case, is to demonstrate the finite-sample uncertainty of our constructed test statistics can be bounded by a control statistics, whose *exact finite-sample distribution*, and thus the corresponding *critical values*, can be simulated. We also show how to get sharper tests (and critical values) through constrained thresholding, studentization, and more sophisciated tail weighting schemes that dampen the effect of restrictions with low effective sample sizes and high uncertainty.

We demonstrate in a simulation study the performance of our single-network confidence sets, with network sizses ranging from 100 to $10,000$. Our parametric confidence sets certifies

the sign of the strategic coefficient γ_0 in every replication from $n = 400$ onward. We also show how two semiparametric versions (one with and one without a symmetry assumption on F_ε) of our inference procedures still produce nontrivial sign-revealing confidence sets at reasonable sample sizes.

We also apply our inference procedure to two empirical data sets: one on high-school contact networks ($n = 327$) and another on Twitch gamer networks (n up to $9,498$). In both settings, we find strictly positive (projected) confidence intervals for the strategic coefficients at 95%-level.

Our paper contributes mostly directly to the econometric literature on strategic network formation, with the following two highlights:

The first highlight is *pathwise identification* in a single large network setting, where the formulation is as novel as the restrictions. Identification analysis normally starts from a deterministic feature of the observable distribution: a population probability, moment, or probability limit that the data can recover and the model restricts. Here no such objects need exist, and thus the valid *identified set* can still be *random even in the limit*.[4] What becomes deterministic in our setting is the limit of an exogenous statistic, which encodes the exogeneity assumption in the model as the identifying leverage. We call the resulting notion *pathwise identification*, and we know of no precedent for it in the network formation literature.

The second highlight is *finite-sample inference*, which we regard as the central contribution of this paper to the network econometrics literature: to our knowledge these are the first computationally and theoretically tractable, finite-sample valid confidence sets for the structural parameters of a strategic network formation game observed as a single large network under unknown equilibrium selection, with no restrictions on density or on the dependence the game induces, and with no equilibrium simulation. The tractability is structural rather than incidental. Testing a candidate parameter requires no equilibrium solve, no search over completions of the observed network, and—because our restrictions are indexed by dyads rather than by subnetwork configurations—no graph isomorphism matching. What the criterion needs are empirical averages of indicator variables over the dyads falling in cells of exogenous characteristics, which is counting. The computational cost therefore roughly

[4]The randomness at issue here is not that of a confidence set. A confidence set is random in any finite sample simply because it is a function of the data; that randomness is standard, well understood, and disappears in the usual asymptotics as the set settles down around a population object. What is unusual here is that randomness may *survive* the large-n limit: absent weak-dependence conditions the observable envelopes need not converge along the realized network sequence, so the limiting identified set is itself a random object rather than a feature of the population distribution.

scales with the number of dyads, $n(n-1)/2 = O(n^2)$, rather than with the size $2^{n(n-1)/2}$ of the graph space. The practical consequence is that network sizes out of reach for simulation-based methods are routine here: the simulation study runs to $n = 10{,}000$ and the empirical application to $n = 9{,}498$, each returning a certified confidence set in manageable time.

Literature Review

The closest papers in the strategic network formation literature are de Paula, Richards-Shubik, and Tamer (2018) and Menzel (2026). de Paula, Richards-Shubik, and Tamer (2018) study partial identification of preferences in a *single large network* formed under pairwise stability, with non-transferable utility and complete information. They impose two restrictions to make the payoff-relevant environment finite: utility depends on the network only out to a bounded distance, and all agents have a bounded number of links. Each agent is then classified by her *network type* (defined jointly over the graph of her neighborhood and the covariates of the nodes in it), and parameters are constrained by the observed distribution of network types. Their computation is carried out through a series of quadratic programs in a continuum-of-agents formulation rather than at a finite n, and matching network types requires deciding joint covariate-and-graph isomorphism, which is computationally hard in large finite networks. We give up sharpness and the local combinatorics, using only the threshold-crossing structure of a single dyad. In exchange we impose no bound on degree, none on interaction depth, and no restriction on density. Menzel (2026), who studies a similar model in a single large network, also relies on dyad-level link frequencies to deliver identification and estimation. The first difference is that, while endogenous covariates are allowed in Menzel (2026), they must take a restrictive "node-incident" form that rules out "fundamentally dyadic" network statistics such as common friends. The methodology is also very different from ours: he derives large-network asymptotic approximations, obtaining a Poisson likelihood limiting model under a Type I upper tail condition on the idiosyncratic error and unique edge response condition as well as equilibrium selection. The results are also asymptotic approximations, where ours are exact at the observed finite n.

Some related work consider similar strategic network formation models, but focus on the many-network setting for identification and inference. Sheng (2020) studies the same model with a complementary strategy in a many network settting, who derive identifying restrictions by checking pairwise stablility conditions on subnetwork structures and resolve multiple equilibria using the 0-1 bounds a la Tamer (2003). Her approach also requires a network isomorphism algorithm that selects matches on specific subnetwork configurations where the bounds are tractable to compute, and the inference theory is built for many independent (and typically small) networks, with asymptotics established in the number

of independent networks. Gualdani (2021) also works with many observed networks, in a complete-information game with directed links and a spillover in the number of agents linking to the same target. Her device is a decomposition of the formation game into *local games* that are in equilibrium if and only if the whole game is, which cuts the number of moment inequalities characterizing the sharp identified set and makes the integrals entering them computable.

There are also other routes that resolve the equilibrium multiplicity issue rather than bracketing it. Miyauchi (2016) exploits nonnegative externalities and the resulting lattice structure to bracket moments between extremal equilibria, computed by simulation. Mele (2017) and Badev (2021) model equilibrium selection directly, treating the observed network as a draw from the stationary distribution of a myopic adjustment process, so that estimation on a single network satisfies mixing requirements. A separate branch weakens the informational environment instead: with private information, realized links are conditionally independent given types, which severs the global dependence and permits two-step asymptotically normal estimation on one network (Leung, 2015; Ridder and Sheng, 2025; Comola and Dekel, 2026; Wu, 2024). Hence, the model, methodology, and results obtained are quite different from ours.

In a companion paper, we further combine bounding-by-c with subnetwork differencing to accommodate strategic interdependence *and* unobserved individual fixed effects (Gao, Li, and Xu, 2026), but the inclusion and differencing of fixed effects imply that no node-level covariates can be included.

Our paper also relates to the larger network formation literature. Complementary to the strategic network formation literature where our paper belongs, the dyadic network formation literature studies models with agent unobserved heterogeneity but *no* link interdependence: see, e.g., Graham (2017); Jochmans (2018); Dzemski (2019); Gao (2020); Zeleneev (2026), as well as work that focuses on the nontransferable-utility case (Gao et al., 2023; Li et al., 2026; Marshall, 2026). That conditional independence is exactly what makes those models tractable, and it is exactly what link interdependence invalidates: our whole difficulty is that X_{ij} is a function of the entire realized network.

There have also been theoretical advances on *inference in a single network setting*, which mostly focuses on deriving valid asymptotic results under suitable conditions on network dependence: ψ-dependence decaying in network distance (Kojevnikov et al., 2021), approximate neighborhood interference (Leung, 2022), exchangeable-array structure (Davezies et al., 2021; Graham, 2020), or a density regime that bounds the strength of interdependence (Leung, 2019; Menzel, 2026; Chandrasekhar and Jackson, 2025). The most recent advance that

delivers a central limit theorem under a network formation model with explicit strategic interdependence is Leung and Moon (2026), which requires branching-process subcriticality of the spillovers *and* a sufficiently decentralized selection mechanism. We impose no counterpart to any of these and provide finite-sample inference results instead of asymptotic ones. That said, the corresponding cost is that we obtain coverage rather than an asymptotic distribution, and hence conservatism. The two approaches are therefore complementary rather than competing. Where the conditions of Leung and Moon (2026) do hold, their central limit theorem could in principle be used to calibrate critical values to the actual sampling distribution of our dyad-level statistics, buying sharpness at the informative margin in exchange for the assumption-free validity we insist on here. We leave that combination to future work.

More generally, our paper contributes to the literature on *incomplete models and partially identifying inequalities.* Our restrictions are moment inequalities, and their population treatment is the standard one for incomplete models: bracket outcome probabilities between some- and every-equilibrium envelopes (Tamer, 2003; Beresteanu et al., 2011; Galichon and Henry, 2011) or model selection directly (Bajari et al., 2010), as surveyed by Aradillas-López (2020). What differs here is that we never evaluate nor invert the equilibrium correspondence, so we make no claim to sharpness, which is plausibly intractable in a single large network.[5] The inference we build is not the asymptotic inequality inference of Andrews and Shi (2013) and Chernozhukov et al. (2013), which would require exactly the dependence theory we avoid. Methodologically, the "bounding-by-c" device descends from the partial stationarity approach of Gao and Wang (2026) for nonlinear dynamic panel models, and the aggregation of inequalities by logical and monotonicity operations relates to Gao, Li, and Xu (2023).

Lastly, our finite-sample inference procedure relates to finite-sample randomization and Monte Carlo tests (Besag and Clifford, 1989). What has kept this tradition out of strategic network formation is that the null distribution of any statistic touching X_{ij} depends on the intractable equilibrium map and the unknown selection mechanism, so it is neither simulable nor known by symmetry. This is solved by the bounding-by-c technique, after which the finite-sample inference validity follows with technical ease. In econometrics, the closest work is Rosen and Ura (2025), who obtains finite-sample conditional inference with simulable critical values. However, their model setup is the semiparametric binary choice model underlying the maximum score estimator (Manski, 1975, 1985, 1987), which is a cross-sectionally i.i.d. single-agent model, with no strategic equilibrium, multiplicity or dependence issues. In network econometrics, exact inference mostly takes the form of *sharp-null* tests, e.g., Graham and Pelican (2020); Pelican and Graham (2022), the two-sample test of Auerbach (2022),

[5] See more discussion about sharpness after Theorem 1

and design-based causal inference with network interference (Athey et al., 2018; Puelz et al., 2022). A recent paper by Kasy et al. (2026) also obtains exact finite-sample permutation tests and conservative confidence intervals a network formation setting. However, Kasy, Linos, and Mobasseri (2026) focuses on a causal inference framework with known randomization of the initial network, and requires observation of the network structure over a panel of at least two periods; in addition their object of interest is on causal effects induced by an experiment rather than structural network formation parameters. Also related are Kaido and Zhang (2025) and Li and Henry (2026), who obtain finite-sample valid, selection-robust confidence sets for incomplete models using *many independent* observations with parametric latent variables (see also Epstein, Kaido, and Seo, 2016). A very recent paper from the statistics literature, Yanchenko et al. (2026), provides finite-sample model selection test on statistical network formation models (such as random graph and stochastic block/communty models) that includes neither exogenous or endogenous covariates. Our strategic network formation setting is thus out of the scope of all these papers.

The remainder of the paper is organized as follows. Section 2 presents the model setup, the key realization-wise bounding-by-c inequalities, and the identification analysis. Section 3 develops the finite-sample inference under both parametric and semiparametric settings. Section 4 reports the simulation study and Section 5 the empirical application. All proofs are available in Appendix A.

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, 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. 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, where a link between agents i and j exists if and only if the latent surplus from the link is non-negative:

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

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 function w_n, symmetric in its two arguments, which may encode homophily, level, and interaction effects: see leading examples of Z_{ij} below;

- X_{ij} is a vector of potentially endogenous covariates, which may involve other links Y_{hk} with $(h,k) \neq (i,j)$: we discuss X_{ij} in more detail below;

- ε_{ij} is an idiosyncratic pairwise shock.

We allow the covariate function 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.

The structural parameters of interest are $\theta_0 := (\beta_0, \gamma_0) \in \Theta$. Throughout this paper, we use the shorthand notation

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

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

The exogenous dyadic covariates Z_{ij} can include a constant 1 (under a proper location normalization in the distribution of ϵ), homophily effects in the form of component-wise absolute difference $|Z_i - Z_j|$ for continuous variables or $\mathbf{1}\{Z_i = Z_j\}$ for discrete variables, as well as level effects in the form of $Z_i + Z_j$.[6]

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. Specifically, we allow:

$$X_{ij} = \phi_{ij}\left(Y_{-ij}, Z\right) \tag{3}$$

where ϕ_{ij} is a known function mapping the ij-excluded[7] realized network Y_{-ij} and exogenous

[6]The ability of including level effects of the form $Z_i + Z_j$ is one of the key flexibility in network formation models without unobserved individual fixed effects: in the dyadic literature with additive degree heterogeneity (Graham, 2017; Dzemski, 2019; Gao, 2020), the level component is collinear with the fixed-effect profile and is annihilated by every device that eliminates it, so only difference-type effects can be included.

[7]This restriction is only required to endow model (1) with the *economic* interpretation of "pairwise stability", which is formulated on the utility difference of a *counterfactual* comparison between linking and not linking. Note that ϕ_{ij} above can be defined as

$$\phi_{ij}\left(Y_{-ij}, Z\right) = \tilde{\phi}_{ij}\left(\left(Y_{-ij}, 1\right), Z\right) - \tilde{\phi}_{ij}\left(\left(Y_{-ij}, 0\right), Z\right),$$

where (Y_{-ij}, y_{ij}) denotes the two networks under the two possible own-link states $y_{ij} = 1$ and $y_{ij} = 0$. Here, $\tilde{\phi}_{ij}$ is a function that can depend on the whole network structure, including own link ij: for example, the eigenvetor centrality of ij. That said, even if X_{ij} does depend on the realized Y_{ij}, all the econometric

characteristics $Z = (Z_1, \ldots, Z_n)$ to a vector of dyadic covariates. We allow ϕ_{ij} to be both locally and globally defined (subject to the subtlety in Footnote 7).

One canonical local network statistic is the number of common friends of i and j, together with its normalized version,

$$\mathrm{CF}_{ij} := \sum_{k \neq i,j} Y_{ik} Y_{jk}, \qquad \overline{\mathrm{CF}}_{ij} := \frac{\mathrm{CF}_{ij}}{n-2} \in [0,1], \tag{4}$$

capturing transitivity: the surplus of link ij may increase with the number of shared neighbors. The normalized version is the friends-in-common statistic of Sheng (2020), and a nonnegative coefficient on it delivers the nonnegative externalities exploited by Miyauchi (2016). It is the statistic we focus on in simulation and the empirical application.

Other leading choices of local network statistics are accommodated without modification: for example, the Jaccard overlap index $J_{ij} := \frac{\mathrm{CF}_{ij}}{\sum_{k \neq i,j} \mathbf{1}\{Y_{ik} + Y_{jk} \geq 1\}}$ (with $J_{ij} := 0$ when the denominator vanishes), which measures overlap *relative* to the union of the two neighborhoods and is a nonlinear, *non-monotone* functional of the network; second-order friends in the form of $\overline{\mathrm{SF}}_{ij} := \frac{1}{n-2} \sum_{k \neq i,j} (Y_{ik} + Y_{jk}) \in [0,2]$, 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 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 (subject to the subtlety in Footnote 7). Such globally defined X_{ij} are usually highly nonlinear and implicit aggregation of links from the whole network Y, for which there exist no closed-form representations.

Since X_{ij} may depend on the realized network Y, while each link in Y is determined by equation (1) simultaneously, model (1) can be interpreted as the equilibrium network in a strategic network formation game under transferable utilities with pairwise stability as the equilibrium solution concept (cf. Jackson and Wolinsky, 1996; Sheng, 2020). In general, the game may have multiple equilibria for a given realization of primitives (Z, ε), where $\varepsilon = (\varepsilon_{ij})_{i<j}$ collects all the idiosyncratic shocks. We represent the realized network abstractly as:

$$Y = g(Z, \varepsilon; \theta_0, \xi) \tag{5}$$

inference results in our paper still apply (under the subsequently stated assumptions), with the only caveat being that we should no longer interpret (1) as a literal pairwise stability condition.

where g incorporates both the equilibrium correspondence and an (arbitrary and possibly unknown) equilibrium selection mechanism that picks a single equilibrium for each realization of the primitives, potentially depending on an unrestricted random element ξ that drives equilibrium selection.

The equilibrium mapping g is generally intractable to characterize. Even for simple specifications of ϕ_{ij}, the fixed-point nature of the equilibrium creates a complex interdependence structure, which is exacerbated by the discrete combinatorial nature of the graph space. A globally defined ϕ_{ij} would induce yet another layer of complexity and global dependence. As will be shown below, our identification approach does not require characterization, evaluation, or computation of g: we extract identifying information using monotonicity restrictions that hold regardless of the complexity of the equilibrium correspondence and any equilibrium selection mechanisms.

We maintain the following assumptions throughout.

Assumption 1 (Model Specification)**.** *A single network $Y := (Y_{ij})_{i<j}$ is observed along with individual covariates $Z := (Z_i)_{i=1}^n$, and (Y, Z) satisfies model (1) for some unknown parameter θ_0 and unobserved shock realization $\varepsilon = (\varepsilon_{ij})_{i<j}$.*

Assumption 2 (Random Sampling)**.** *The individual-level exogenous characteristics Z_i are independently and identically distributed across agents $i = 1, \ldots, n$.*

Assumption 3 (IID Pairwise Errors)**.** *The idiosyncratic shocks ε_{ij} are independently and identically distributed across all pairs (i, j) with $i < j$, with common CDF F_ε.*

Assumption 4 (Exogeneity/Independence)**.** *The vector of exogenous characteristics Z is independent of the idiosyncratic shocks ε.*

Assumptions 1-4 are standard in the strategic network formation literature. Assumption 1 essentially assumes that a pairwise stable network exists and the observed network is pairwise stable. Assumption 2 asserts random sampling of exogenous covariates across individuals. Assumption 3 imposes the idiosyncrasy of link-level surplus shocks but leaves their common distribution F_ε unrestricted for now: our key identifying strategy applies to both parametric and semiparametric settings, and we will introduce a parametric assumption on F_ε when it is explicitly required. Assumption 4 is an exogeneity condition on the observables Z and the pairwise error ε. Importantly, we do *not* assume that X and ε are independent: since X is a function of the equilibrium network it is endogenous and generally correlated with the shocks ε. This assumption also formalizes the naming convention of calling Z the exogenous covariates while calling X the endogenous covariates.

Assumption 1 also stresses that the data environment we consider is a *single network* observed once. The alternative environment of *many independent networks* (villages, schools, classrooms, markets) is simpler in the sense that conditional probabilities of observable events given the agents' exogenous characteristics are population moments that are directly identified and consistently estimable across networks with no restriction on within-network dependence. Our restrictions also apply to that case, and see Remark 3 for more discussion. That said, the single network environment is what our procedures are mainly designed for and the one we focus on in the main text.

2.2 The "Bounding-by-c" Event Inequalities

This subsection develops the identification strategy at the level of *events*: every statement holds realization by realization, i.e., for every value of the primitives, every equilibrium consistent with them, and every selection among equilibria. No expectation, limit, or probability model enters. These realization-wise inequalities are the cleanest foundation for the identifying restrictions of Section 2.3 as well as the finite-sample inference theory of Section 3.

Under model (1), each link satisfies the threshold-crossing representation $Y_{ij} = \mathbf{1}\{\varepsilon_{ij} \leq \delta_{ij}\}$ with $\delta_{ij} = \delta_{ij}(\theta_0) = Z'_{ij}\beta_0 + X'_{ij}\gamma_0$. The essential difficulty is that the threshold δ_{ij} contains the endogenous covariate X_{ij}, which is co-determined with the entire shock vector ε through the equilibrium: the event $\{\varepsilon_{ij} \leq \delta_{ij}\}$ couples the shock to an endogenous, equilibrium-determined random threshold, and its probability depends on the intractable equilibrium mapping g and on the unknown selection mechanism.

The "bounding-by-c" idea is to trade the endogenous threshold for a deterministic one, which we now explain. Fix any constant $c \in \mathbb{R}$ and consider the event

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

Given that $\{Y_{ij} = 1\} = \{\varepsilon_{ij} \leq \delta_{ij}\}$, together with the bounding-by-c 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 number c. Similarly, a flipped event

$$Y_{ij} = 0 \quad \text{and} \quad \delta_{ij} \geq c$$

implies $\varepsilon_{ij} > \delta_{ij} \geq c$ and thus $\varepsilon_{ij} > c$. Formally, for any $c \in \mathbb{R}$, we have

$$Y_{ij}\mathbf{1}\left\{\delta_{ij} \leq c\right\} \;\leq\; \mathbf{1}\left\{\varepsilon_{ij} \leq c\right\} \leq 1 - (1 - Y_{ij})\,\mathbf{1}\left\{\delta_{ij} \geq c\right\}. \tag{6}$$

Three features of (6) deserve emphasis, because they carry the entire methodological weight of the paper.

First, the event inequalities (6) 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 below are robust to equilibrium multiplicity and arbitrary, unknown selection.

Second, (6) *can be aggregated* under any nonnegative weights or more generally, any weakly increasing transformation: averaging over the dyads of a covariate cell the cornerstone of the inference procedures of Section 3), summing signed certificates across the links of a configuration (Remark 1), and taking conditional expectations in any environment where they are identified (Remark 3). In every such aggregate the endogeneity of X_{ij} is immaterial: no conditional distribution of X_{ij} given Z is ever modeled, estimated, or simulated—the endogenous covariate enters only through the observable indicator $\mathbf{1}\{\delta_{ij}(\theta) \leq c\}$.

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 $F_\varepsilon(c)$. Hence, sweeping c over $\mathbb{R}$ is effectively a way to trace out an envelope for the entire error CDF $F_\varepsilon(c)$. 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 types $Z_D := (Z_i, Z_j)$, which we will explain in the next subsection.

Above we presented bounding-by-c inequalities for dyad ij under transferable utility, which is the case we carry through the paper for expositional simplicity. The technique is tied to neither restriction: it applies to general subnetwork configurations and to nontransferable utility as well, as the next two remarks illustrate. While the identifying restrictions of Section 2.3 and the inference procedures of Section 3 are plausibly adaptable to each setting, we leave these extensions to future work.

Remark 1 (General Subnetwork Configurations). To illustrate, consider a triad ijk. Bounding by c applies to any signed combination of its links, giving inequalities such as

$$\begin{aligned} Y_{ij}Y_{ik}\mathbf{1}\{\delta_{ij} + \delta_{ik} \leq c\} &\leq \mathbf{1}\{\varepsilon_{ij} + \varepsilon_{ik} \leq c\} \\ Y_{ij}(1 - Y_{ik})\mathbf{1}\{\delta_{ij} - \delta_{ik} \leq c\} &\leq \mathbf{1}\{\varepsilon_{ij} - \varepsilon_{ik} \leq c\} \\ Y_{ij}Y_{ik}(1 - Y_{jk})\mathbf{1}\{\delta_{ij} + \delta_{ik} - \delta_{jk} \leq c\} &\leq \mathbf{1}\{\varepsilon_{ij} + \varepsilon_{ik} - \varepsilon_{jk} \leq c\}. \end{aligned}$$

and their "flipped" counterparts. In each case the right-hand side is free of endogenous covariates: it is a threshold event in a signed sum of shocks, whose distribution is a known signed

convolution of F_ε. These are illustrations rather than an exhaustive list, and analogous inequalities can be derived for any subnetwork configuration. Unlike subnetwork approaches built on stability conditions, no matching of graph isomorphism is required: each configuration contributes a closed-form counting restriction, and the researcher may use as few or as many as the data and computational resources support.

Remark 2 (Non-Transferable Utility). Relatedly, the technique applies to the nontransferable-utility model

$$Y_{ij} = \mathbf{1}\left\{Z'_{ij}\beta_0 + X'_{ij}\gamma_0 \geq \varepsilon_{ij}\right\}\mathbf{1}\left\{Z'_{ji}\beta_0 + X'_{ji}\gamma_0 \geq \varepsilon_{ji}\right\} \tag{7}$$

in which $(Z_{ij}, X_{ij}, \varepsilon_{ij})$ may be *asymmetric* across ij and ji, so that the undirected link $Y_{ij} \equiv Y_{ji}$ requires bilateral consent. The following inequalities then hold:

$$\begin{aligned}
Y_{ij}\mathbf{1}\{\delta_{ij} \leq c\} &\leq \mathbf{1}\{\varepsilon_{ij} \leq c\} \\
Y_{ij}\mathbf{1}\{\delta_{ji} \leq c\} &\leq \mathbf{1}\{\varepsilon_{ji} \leq c\} \\
Y_{ij}\mathbf{1}\{\delta_{ij} \leq c_1\}\mathbf{1}\{\delta_{ji} \leq c_2\} &\leq \mathbf{1}\{\varepsilon_{ij} \leq c_1\}\mathbf{1}\{\varepsilon_{ji} \leq c_2\} \\
(1 - Y_{ij})\mathbf{1}\{\delta_{ij} > c\}\mathbf{1}\{\delta_{ji} > c\} &\leq 1 - \mathbf{1}\{\varepsilon_{ij} \leq c\}\mathbf{1}\{\varepsilon_{ji} \leq c\}
\end{aligned}$$

and many more can be derived from other subnetwork structures. The contrast between the first three inequalities and the last is the substantive feature of this environment: a formed link certifies both directional shocks and is therefore attributable to each margin separately, whereas an absent link certifies only that at least one margin refused, and so constrains the two jointly.

2.3 Identifying Restrictions under Large-Network Asymptotics

"Standard identification analysis" usually requires converting inequalities on random variables into deterministic population (or asymptotic) restrictions after finite-sample randomness of the model is averaged out. To do so in a single large network setting, one might expect that we would need delicate weak-dependence conditions. The main message of this subsection is that no such conditions are needed for the *validity* of our identifying restrictions. This is because the bounding-by-c inequalities hold realization by realization on the observed network, and the only unknown they sandwich is the distribution of a *purely exogenous* quantity, whose empirical counterpart converges in the large-network limit by classical limit theory for exchangeable arrays, regardless of what the equilibrium correspondence and the selection mechanism do.

Formally, we consider the following single large-netwok asymptotics. We place all prim-

itives on a single probability space: an infinite i.i.d. sequence $(Z_i)_{i\geq 1}$, an infinite array $(\varepsilon_{ij})_{ij}$ of i.i.d. shocks independent of Z, an auxiliary randomization variable U_{sel} independent of (Z, ε) (allowing randomized equilibrium selection), and, for each n, a realized network $Y^{(n)} = g_n(Z^{(n)}, \varepsilon^{(n)}, U_{\text{sel}}; \theta_0)$ formed from the first n agents by the (n-dependent) equilibrium-and-selection mapping, as in (5). We write $\mathbb{P}$ for the joint law of $\big((Z_i)_{i\geq 1}, (\varepsilon_{ij})_{ij}, U_{\text{sel}}\big)$. The equilibrium correspondence and the selection mechanism embodied in g_n may change arbitrarily with n and may depend on all primitives.

We now explain how we derive the large-network identifying restrictions.

The Finite-Sample Sandwich

We start by working with finite-sample conditional average of the bounding-by-c inequalities derived in Section 2.2.

We first define a discrete dyad-type 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)**.** Let $(\mathcal{Z}_{D,1}, ..., \mathcal{Z}_{D,\kappa})$ be an arbitrary finite, *symmetric*[8] partition of $\text{Supp}\,(Z_i, Z_j) \equiv \text{Supp}\,(Z_i)^2$ such that $\mathbb{P}\left((Z_i, Z_j) \in \mathcal{Z}_{D,\kappa}\right) > 0$ for each $k = 1, ..., \kappa$. We define the discretized dyad type $Z_{D,ij}$ as

$$Z_{D,ij} := z_{D,k} \quad \text{if } (Z_i, Z_j) \in \mathcal{Z}_{D,k}$$

where $z_{D,k}$ are κ arbitrary cell labels. We write $\mathcal{Z}_D := \{z_{D,1}, ..., z_{D,\kappa}\}$ 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, which will result in *no loss of validity* as shown below.[9] When Z_i has finite support, one can always take the partition to be each support point in $\text{Supp}\,(Z_i, Z_j)$. 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.

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

[9] We work throughout with the discretized dyad type because it yields the cleanest conditional-mean formulation: the cells partition the dyads exactly, so the conditioning is exact and distinct cells hold disjoint sets of shocks. What discretization costs is informational coarsening, not validity: the restrictions hold cell by cell for any partition, so a coarser partition simply delivers fewer of them. For continuous Z_i, we could in principle use a finer conditioning device by multiplying the pointwise inequalities with a nonnegative kernel, at the cost of additional regularity conditions for the limiting statements. In finite samples, however, the trade-off often runs the other way, and coarsening can be actively desirable. Cell counts govern how tightly the exogenous middle term concentrates, so in sparse networks, where fine cells hold few dyads, a coarser partition buys sharper control of that term and hence a smaller critical value (See Section 3 on critical values in finite-sample inference).

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

$$\hat{p}_L^{(n)}(z_D, c; \theta) := \frac{\sum_{ij} Y_{ij} \mathbf{1}\{\delta_{ij}(\theta) \leq c\} \mathbf{1}\{Z_{D,ij} = z_D\}}{\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\}}, \tag{8}$$

$$\hat{p}_U^{(n)}(z_D, c; \theta) := 1 - \frac{\sum_{ij} (1 - Y_{ij}) \mathbf{1}\{\delta_{ij}(\theta) \geq c\} \mathbf{1}\{Z_{D,ij} = z_D\}}{\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\}}, \tag{9}$$

together with the *exogenous middle term*

$$\hat{F}_n(c; z_D) := \frac{\sum_{ij} \mathbf{1}\{\varepsilon_{ij} \leq c\} \mathbf{1}\{Z_{D,ij} = z_D\}}{\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\}}. \tag{10}$$

All three objects are defined on the event that $\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\} \geq 1$, which occurs for all large n almost surely since $\mathbb{P}(Z_{D,ij} = z_D) = \mathbb{P}((Z_i, Z_j) \in \mathcal{Z}_k) > 0$. 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 $Z_{D,ij}$.[10] The frequencies $\hat{p}_L^{(n)}$ and $\hat{p}_U^{(n)}$ are computable from observed data (Y, X, Z),[11] while the middle term $\hat{F}_n$ involves the unobserved shocks and serves purely as a theoretical bridge and is never computed.

The per-dyad inclusions (6) hold *pointwise*—for every realization of the primitives, every equilibrium, and every selection mechanism. Averaging within a cell therefore yields the *finite-network oracle sandwich*: for every n and *every realization*,

$$\hat{p}_L^{(n)}(z_D, c; \theta_0) \;\leq\; \hat{F}_n(c; z_D) \;\leq\; \hat{p}_U^{(n)}(z_D, c; \theta_0) \qquad \text{for all } c \in \mathbb{R} \tag{11}$$

at every z_D such that $\sum_{ij} \mathbf{1}\{Z_{D,ij} = z_D\} \geq 1$. The sandwich inequality (11) is stronger than a population moment inequality: it holds *before* any expectation is taken, and under arbitrary dependence in the realized network. All equilibrium complications are confined to the two observable outer terms, while the middle term contains only the independent Z and ε.

[10] We are therefore not "discretizing the model" in any way. If Z_i is continuous in the model, then $\delta_{ij}(\theta)$ continues to vary continuously with Z_{ij}, and β remains the coefficient on the raw covariate. The discretized type $Z_{D,ij}$ enters only as the conditioning variable through which dyads are aggregated into the cell averages above.

[11] The upper frequency (9) scans the strict index event $\{\delta_{ij}(\theta) > c\}$. The weak version $\{\delta_{ij}(\theta) \geq c\}$ is equally valid—an unformed link certifies $\varepsilon_{ij} > \delta_{ij}(\theta_0)$ strictly, so $\delta_{ij}(\theta_0) \geq c$ still delivers $\varepsilon_{ij} > c$—and is weakly tighter. We maintain the strict convention throughout.

A Uniform Strong LLN for the Exogenous Middle Term

What drives our large-network identification result is the convergence of the middle term $\hat{F}_n(c; z_D)$ as $n \to \infty$.

Lemma 1 (Uniform LLN for the Exogenous Middle Term)**.** *Suppose Assumptions 2–4 hold. Then*

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

The proof (Appendix A) is classical: the numerator and denominator of $\hat{F}_n$ are U-statistics of the jointly exchangeable, dissociated array formed by the i.i.d. types (Z_i) and the i.i.d. dyadic shocks (ε_{ij}), so the strong law of large numbers for exchangeable arrays applies. The equilibrium and the endogenous X_{ij} play no role in this convergence.

Pathwise Identified Set

Combining the oracle sandwich with the uniform strong law fixes the profiling step. Define, for each (z_D, c, θ), the (possibly random) limit frequencies

$$p_L^\infty(z_D, c; \theta) := \limsup_{n \to \infty} \hat{p}_L^{(n)}(z_D, c; \theta), \qquad p_U^\infty(z_D, c; \theta) := \liminf_{n \to \infty} \hat{p}_U^{(n)}(z_D, c; \theta), \tag{12}$$

which always exist. Crucially, the inequality $\limsup_n \hat{p}_L^{(n)} \leq \liminf_n \hat{p}_U^{(n)}$ does *not* follow from (11) alone, but it does follow once the middle term is known to *converge*: by Lemma 1, almost surely,

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

This yields the population version of single-network validity.

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

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

Consequently, almost surely:

- *(i) Semiparametric identified set:* $\theta_0 \in \Theta_I^\infty$*, where*

$$\Theta_I^\infty := \left\{ \theta : \sup_{z_D} p_L^\infty(z_D, c; \theta) \leq \inf_{z_D} p_U^\infty(z_D, c; \theta) \textit{ for all } c \in \mathbb{R} \right\}; \tag{13}$$

- *(ii) Parametric identified set: if in addition F_ε belongs to a known parametric family $F_\varepsilon(\,\cdot\,;\theta_\varepsilon)$, then the stacked true parameter $\overline{\theta}_0 = (\theta_0, \theta_{\varepsilon,0})$ satisfies*

$$\sup_{z_D} p_L^\infty(z_D, c; \theta_0) \;\le\; F_\varepsilon(c; \theta_{\varepsilon,0}) \;\le\; \inf_{z_D} p_U^\infty(z_D, c; \theta_0) \qquad \text{for all } c \in \mathbb{R}. \tag{14}$$

We refer to Θ_I^∞ (or its parametric counterpart) as the *pathwise identified set*: it is defined along the realized network sequence.

We now provide some discussion about Theorem 1.

First, we clarify that we use the phrase "identified set" *without* claiming sharpness, and there is a substantive reason to doubt sharp characterizations are feasible in this class. Sharp identified sets in games with multiple equilibria are often characterized by, in effect, solving the model: for each candidate parameter and each realization of the latent primitives, enumerating the full set of equilibrium outcomes and checking 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 et al. (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, which is exactly the motivation behind this paper. The value of the bounding-by-c restrictions is that they extract equilibrium-robust identifying content *without ever solving the game*, and the finite-sample inference of Section 3 is built on their realization-wise structure rather than on sharpness.

Second, Theorem 1 is a *validity* statement rather than a characterization of a deterministic population quantity. Without weak-dependence conditions, $\hat{p}_L^{(n)}$ and $\hat{p}_U^{(n)}$ need not converge: 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*. This thus marks a departure from the standard 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 the model restricts. 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 sample means are in general random variables in the form of conditional expectations given the invariant or exchangeable σ-field rather than constants (cf. Eagleson and Weber, 1978; Davezies et al., 2021), and inference can still proceed conditionally on the realized time-series path. In the same spirit, identification here is articulated *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.

Remark 3 (The Many-Network Environment in One Paragraph). When the researcher observes K independent networks drawn from a common data-generating process, everything above simplifies and the limit frequencies (12) are replaced by ordinary population moments. Taking conditional expectations directly in (6) yields

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

with

$$\begin{aligned} p_L(z_D, c; \theta) &:= \mathbb{E}[Y_{ij}\mathbf{1}\{\delta_{ij}(\theta) \le c\} \mid Z_{D,ij} = z_D], \\ p_U(z_D, c; \theta) &:= 1 - \mathbb{E}[(1 - Y_{ij})\mathbf{1}\{\delta_{ij}(\theta) > c\} \mid Z_{D,ij} = z_D], \end{aligned}$$

based on which we can obtain a many-network analog of Theorem 1. These moments are directly identified and consistently estimable as $K \to \infty$ by the ordinary law of large numbers across independent networks, with *no* restriction on within-network dependence. Here, p_L and p_U are understood as conditional probabilities that are deterministic at each given z_D, which is "standard" in identification analysis.

Remark 4 (Location and Scale Normalization). As standard in network formation (or more generally binary outcome) models, our modelrestricts the pair $(\theta_0, F_\varepsilon)$ only up to location and scale normalization.[12] When F_ε is specified parametrically as, say, normal or logistic distribution, it is without loss of generality to set F_ε as the standard normal or the standard logistic. When F_ε is left nonparametric, we normalize the constant to zero and fix the scale of one nonzero coefficient, say $|\beta_1| = 1$.

Remark 5 (Shape Restrictions on F_ε). Because the middle term of the sandwich is exactly F_ε itself, nonparametric shape restrictions can be layered onto it directly and tractably: one simply requires the two observable envelopes to be consistent with some CDF in the

[12] Adding a constant a to every index and replacing ε by $\varepsilon + a$ leaves the link equation (1) unchanged, and so does multiplying the index and the shock by a common $b > 0$.

restricted class. This is a further advantage of the bounding-by-c approach: the argument delivers the error CDF as the very object being sandwiched, so global shape restrictions on F_ε can be naturally attached to this middle term. Symmetry about an unknown center and log-concavity are the leading examples, and both are satisfied by most parametric families used in practice. Each such restriction weakly tightens the set in exchange for a stronger maintained assumption, interpolating between the semiparametric criterion, which profiles F_ε out entirely, and the parametric one, which fixes it as an assumption. See Appendix B for more detail about how symmetry restrictions can be formally incorporated without further functional form assumptions.

3 Finite-Sample Inference in a Single Network

The pathwise identification in Section 2.3 is a large-network result along a coupled sequence of networks as $n \to \infty$. This section develops the finite-sample counterpart, which leads to a valid inference procedure that constitutes the paper's central contribution.

3.1 Semiparametric Confidence Sets

We start with the semiparametric case, which turns out easier to present and helps convey the conceptual core of our inference approach.

The semiparametric identifying restriction (13) in Theorem 1 can be encoded by the extremum criterion

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

which satisfies $Q^\infty(\theta_0) = 0$ almost surely. We define its finite-sample analogue by

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

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

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

for a pre-chosen $\underline{m} \geq 1$.[13] Also, even though $\sup_{c\in\mathbb{R}}$ appears to be a supremum taken over

[13] Taking $\underline{m} = 1$ is the minimal requirement for the conditional averages $\hat{p}_L^{(n)}$, $\hat{p}_U^{(n)}$ and $\hat{F}_n$ to be well-defined. Any larger floor is equally valid, and is often preferable in practice. A larger $\underline{m}$ keeps only cells

the whole real line, in finite sample it is actually computable with finitely many evaluations only. 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 exactly at the finitely many realized index values in $\{\delta_{ij}(\theta) : i < j\}$. 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"

$$R_n := \sup_{c \in \mathbb{R}} \Big[\max_{z_D \in \hat{\mathcal{Z}}_D} \hat{F}_n(c; z_D) - \min_{z_D \in \hat{\mathcal{Z}}_D} \hat{F}_n(c; z_D) \Big], \tag{16}$$

which dominates $\widehat{Q}_n(\theta_0)$ as shown in the following lemma.

Lemma 2 (Semiparametric Uncertainty Control)**.** *For every n,*

$$\widehat{Q}_n(\theta_0) \;\leq\; R_n.$$

At a null hypothesis θ_0, the statistic $\widehat{Q}_n(\theta_0)$ on the left is observable and may carry arbitrary network endogeneity and dependence, while the oracle 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 for the truth θ_0.

A key, and somewhat surprising, property of the control statistic R_n is that its conditional distribution given Z can be simulated *without the knowledge of* F_ε.

We describe our proposed inference procedure in the following. For $b = 1, \ldots, B$ with B sufficiently large, draw $\hat{m}(z_D)$ i.i.d. uniform random variables on $[0, 1]$ for each $z_D \in \hat{\mathcal{Z}}_D$, and independently across b and across cells. Let $G_{z_D}^{(b)}$ be the empirical distribution function of cell z_D, and set

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

Let $q_{n,1-\alpha}^R(Z)$ be the $(1-\alpha)$-th quantile[14] of $R^{(b)}$, which we will use as the critical value. We

with enough dyads to carry little finite-sample uncertainty, which lowers the critical value. The effect on the confidence set is a genuine trade-off rather than a free improvement: both $\widehat{Q}_n$ and the critical value are computed over the same family $\hat{\mathcal{Z}}_D$, so raising $\underline{m}$ discards restrictions and lowers the test statistic as well. In sparse networks with many thin cells the first effect typically dominates, which is why we impose a moderate floor rather than $\underline{m} = 1$. The validity of using a pre-chosen floor $\underline{m}$ follows again from the realization-wise validity of the bounding-by-c inequalities.

[14] In fact, we can deliver *exact finite-B coverage guarantee* by using the critical value $\hat{q}_{n,1-\alpha}^R(Z)$, defined as the $\lceil (1-\alpha)(B+1) \rceil$-th smallest value of $R^{(1)}, \ldots, R^{(B)}$. Such choice of $\hat{q}_{n,1-\alpha}^R(Z)$ would guarantee that randomness from the finite-B simulation is also properly accounted for in the confidence set construction. That said, since B can be made arbitrarily large, in the main text we abstract away from finite-B sampling error.

then construct the semiparametric confidence set as

$$\widehat{\mathcal{C}}_n^{\text{semi}}(1-\alpha) := \big\{\theta : \ \widehat{Q}_n(\theta) \le q_{n,1-\alpha}^R(Z)\big\}. \tag{17}$$

Note that the inference procedure uses only the cell sizes $\hat{m}(z_D)$, *without* the need of drawing ε from the unknown distribution F_ε or simulating any network formation process.

Theorem 2 (Semiparametric Confidence Set)**.** *Suppose Assumptions 1–4 hold. Then, for every n,*

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

and hence the coverage guarantee also holds unconditionally.

The proof is in Appendix A, where the key idea is to exploit the probability integral transform (and its discrete analog). Conditional on Z, different dyad cells $z_D \in \hat{\mathcal{Z}}_D$ contain disjoint sets of i.i.d. dyadic shocks (Assumption 3). When F_ε is continuous, setting $U_{ij} := F_\varepsilon(\varepsilon_{ij})$ produces i.i.d. uniform random variables by the probability integral transform, so $\hat{F}_n(c; z_D) = G_{z_D}(F_\varepsilon(c))$, and thus the conditional distribution of R_n only depends on the per-cell sample size $\hat{m}(z_D)$, which can be computed by the $R^{(b)}$ simulation. When F_ε has atoms, the discretized probability integral transform continues to work in delivering validity, but would make the test more conservative than it is in the case with a continuous F_ε.

In fact, a closed-form critical value can be derived using an analytic bound on the tail probability of R_n:

Proposition 1 (Closed-Form Critical Value)**.** *Suppose Assumptions 1–4 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 a solution[15] *to*

$$\sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-\hat{m}(z_D)\hat{t}_n^2/2} = \alpha. \tag{18}$$

Then the confidence set $\{\theta : \widehat{Q}_n(\theta) \le \hat{t}_n(\alpha; Z)\}$ has conditional coverage at least $1-\alpha$.[16]

[15] A solution must exist since $\sum_{z_D \in \hat{\mathcal{Z}}_D} 2e^{-\hat{m}(z_D)\hat{t}_n^2/2}$ decreases continuously and strictly from $2\#\left(\hat{\mathcal{Z}}_D\right)$ to 0 as $\hat{t}_n$ runs over $[0, \infty)$, so the solution exists and is unique for every $\alpha \in (0,1)$.

[16] One can also work with the more conservative bound $\hat{t}_n = \sqrt{2\log\left(2\#\left(\hat{\mathcal{Z}}_D\right)/\alpha\right) / \min_{z_D \in \hat{\mathcal{Z}}_D} \hat{m}(z_D)}$,

Remark 6 (Convergence Rate of Critical Values). Proposition 1 not only provides us with valid inference procedure, but also offers a clean theoretical about the rate at which the critical value shrinks, since (18) pins down the rate at which $\hat{t}_n(\alpha; Z) \to 0$ as $n \to \infty$. To see this, note that the effective sample size $\hat{m}(z_D)$ in cell z_D grows at the order $O(n^2)$, the same as the number of dyads. In the meanwhile, the solution path $\hat{t}_n(\alpha; Z)$ must balance $1/\hat{m}(z_D) \sim \hat{t}_n^2/2$ as $n \to \infty$ to solve (18) at each n. This implies that the critical value $\hat{t}_n(\alpha; Z)$ shrinks at the *dyadic* rate $1/\sqrt{\hat{m}(z_D)} \sim 1/n$ (up to a log factor), 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 induce a convergence rate of $\sqrt{1/n^2} = 1/n$ after averaging.

3.2 Parametric Confidence Sets

We now consider the parametric case, in which F_ε is known up to a finite-dimensional θ_ε. 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.[17] That said, we keep the notation θ_ε for theoretical generality.

Knowing F_ε allows us to leverage the two-sided bounding-by-c bounds separately, in a similar style as the identifying restrictions in Theorem 1(ii). Specifically, we define the finite-sample test statistic

$$\widehat{Q}_n^{par}(\overline{\theta}) := \sup_{c\in\mathbb{R}} \max\Big\{ \max_{z_D\in\hat{\mathcal{Z}}_D} \hat{p}_L^{(n)}(z_D, c; \theta) - F_\varepsilon(c; \theta_\varepsilon),\ F_\varepsilon(c; \theta_\varepsilon) - \min_{z_D\in\hat{\mathcal{Z}}_D} \hat{p}_U^{(n)}(z_D, c; \theta),\ 0\Big\}. \quad (19)$$

Correspondingly, we define the oracle statistic

$$R_n^{par}(\theta_\varepsilon) := \max_{z_D\in\hat{\mathcal{Z}}_D} \sup_{c\in\mathbb{R}} \left|\hat{F}_n(c; z_D) - F_\varepsilon(c; \theta_\varepsilon)\right| \quad (20)$$

which controls the uncertainty in $\widehat{Q}_n^{par}(\overline{\theta})$ at 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,*

$$\widehat{Q}_n^{par}(\overline{\theta}_0) \ \le\ R_n^{par}(\theta_{\varepsilon,0}).$$

obtained by replacing every $\hat{m}(z_D)$ by the smallest one, which does not require solving the equation $\sum_{z_D\in\hat{\mathcal{Z}}_D} 2e^{-\hat{m}(z_D)\hat{t}_n^2/2} = \alpha$.

[17] When F_ε is set to be standard normal or standard logistic, a constant term should be included in Z_{ij}, and the whole vector of θ_0 should be treated as free-varying parameter.

At each null $\theta_{\varepsilon,0}$, we can again simulate the distribution of $R_n^{par}(\theta_{\varepsilon,0})$ to obtain the critical value.

Specifically, for $b = 1, \ldots, B$ with B sufficiently large, draw a full i.i.d. pair-shock array $\{\varepsilon_{ij}^{(b)}\}_{ij}$ from $F_\varepsilon(\cdot\,;\theta_\varepsilon)$, and then compute $\hat{F}_n^{(b)}(c; z_D)$ and

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

Let $q_{n,1-\alpha}(Z; \theta_\varepsilon)$ be the $(1-\alpha)$-th quantile[18] of $R_n^{par,(b)}(\theta_\varepsilon)$, and construct the parametric confidence set as

$$\widehat{\mathcal{C}}_n^{par}(1-\alpha) := \left\{ \overline{\theta} : \ \widehat{Q}_n^{par}(\overline{\theta}) \leq q_{n,1-\alpha}(Z; \theta_\varepsilon) \right\}.$$

Theorem 3 (Parametric Confidence Set)**.** *Suppose Assumptions 1–4 hold with* $F_\varepsilon = F_\varepsilon(\cdot\,;\theta_{\varepsilon,0})$ *known up to the finite-dimensional parameter* $\theta_{\varepsilon,0}$*. Then*

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

and hence the coverage guarantee also holds unconditionally.

3.3 Improved Confidence Sets with Nondecreasing Aggregation

The previous subsections focus on simple finite-sample inference procedures that are direct "finite-sample analogues" of the identifying restrictions in Theorem 1. That said, they are not the only valid procedures induced by the sandwich inequalities (11). Specifically, since (11) is a collection of *pointwise* inequalities, one for each cell z_D and threshold c, we can first transform each side of the inequalities under an arbitrary nondecreasing transformation that preserves the inequalities, and then construct test statistics based on the transformed inequalities. This allows us to better control and account for the varying degrees of uncertainties across z_D and c values, leading to a less conservative test (and thus tighter confidence set) than the ones constructed above.

For concreteness and simplicity, we focus on the *parametric case*. See Remark 7 below on how similar type of improvements also carry over in the semiparametric case.

At a candidate θ, define

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

[18] As explained in Footnote 14, we can use the finite-B adjusted $\hat{q}_{n,1-\alpha}(Z; \theta_\varepsilon)$, defined as the $\lceil (1-\alpha)(B+1) \rceil$-th smallest value of $R_n^{par,(1)}, \ldots, R_n^{par,(B)}$.

Evaluated at the truth $\overline{\theta}_0$, the sandwich inequalities (11) translate to

$$V^{\pm}(z_D, c; \overline{\theta}_0) \;\leq\; R^{\pm}(z_D, c; \theta_\varepsilon) \qquad \text{for every } (z_D, c, \pm),$$

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 $(z_D, c, \pm)$ combinations.

Conditional on Z and at the truth $\overline{\theta}_0$, let T be any *nondecreasing*[19] *aggregation* function, mapping from the space of V or R into a scalar. Then obviously we have

$$T\left(V\left(\overline{\theta}_0\right)\right) \;\leq\; T\left(R\left(\theta_{\varepsilon,0}\right)\right). \tag{21}$$

For example, the statistics $\widehat{Q}_n^{par}$ and R_n^{par} in Section 3.2 are constructed by taking the aggregation function T to be the maximum function over all $(z_D, c, \pm)$ combinations, in which case (21) specializes to Lemma 3.

The above implies the following generalization of the conditional Monte Carlo procedure from Section 3.2.

Conditional on Z and at any candidate $\overline{\theta}$, let $T_{Z,\overline{\theta}}$ be any *known* nondecreasing aggregation as described above. Here we emphasize that $T_{Z,\overline{\theta}}$ can freely depend on $\left(Z, \overline{\theta}\right)$, since our test will be conditional on Z with $\overline{\theta}$ hypothesized as truth under the null. As before, we simulate B i.i.d. copies $R^{(1)}, \ldots, R^{(B)}$ based on Z and $F_\varepsilon\left(\cdot\,; \theta_\varepsilon\right)$, compute $c_{n,1-\alpha}^{T}(Z; \overline{\theta})$ as the $(1-\alpha)$-th quantile of the aggregated statistics $T_{Z,\overline{\theta}}(R^{(b)})$. Then we construct the T-induced confidence set as

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

whose validity is established below.

Theorem 4 (Valid Parametric Confidence Sets under Nondecreasing Aggregation)**.** *Suppose Assumptions 1–4 hold with $F_\varepsilon = F_\varepsilon(\cdot\,; \theta_{\varepsilon,0})$ known up to the finite-dimensional parameter $\theta_{\varepsilon,0}$. Then*

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

and hence the coverage guarantee also holds unconditionally.

This freedom in the choice of T matters because T controls how uncertainty is aggregated and weighted across $(z_D, c, \pm)$, which affects the distribution of $T(R(\theta_\varepsilon)) - T\big(V(\overline{\theta}_0)\big)$. The maximum function used in Lemma 3 is intuitive from the identifying restrictions, but in finite

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

sample it suffers from the sensitivity of taking maximums: one thin cell at an uninformative threshold can become a "maximum" by chance and affects the test statistic or the critical value drastically. Hence, heuristically, we should aggregate the inequalities in a way that controls the influence of any single $(z_D, c, \pm)$ configuration with very few data points (and thus high uncertainty).

Concretely, we consider the following three "obvious improvements". For the rest of this subsection, we suppress θ_ε from F_ε and write θ instead of $\overline{\theta}$: the results still go through at each candidate θ_ε, but, as discussed earlier, in practice F_ε is often taken to be standard normal or standard logistic, leaving no more free θ_ε. Since the rest of the subsection focuses more on practical implementation, this notational simplification helps better convey the key ideas of the improvements without unnecessary clutter in the notation.

(i) Constrained Thresholding.

First, we show how to systematically trim away thresholds c that carry no identifying information about θ and would only inflate the simulated maximum. At each (θ, Z) we take the *constrained threshold set* $C(z_D; \theta, Z)$ to be the *a priori* range of the candidate index in cell z_D. Formally, writing $\mathcal{X}$ as the support of X,

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

where $\min_{\tilde{z}_D \in z_D} w_n(\tilde{z})' \beta$ is the minimum exogenous index across exogenous covariate values $\tilde{z}$ that is consistent with the discretized cell z_D.[20] The point of restricting to $C(z_D; \theta, Z)$ is that it discards thresholds c that could not have bite.[21] That said, for c outside $C(z_D; \theta, Z)$, the control statistic $R^{\pm}(z_D, c; \theta_\varepsilon)$, which does not involves the index, would still inflate the simulated maximum and thus the associated critical value. Hence, for finite-sample performance it is strictly better to throw away c outside $C(z_D; \theta, Z)$. In many practical cases, $\mathcal{X}$ is a compact set, so the restriction to $C(z_D; \theta, Z)$ can be quite substantial.[22]

Given the threshold set $C(z_D; \theta, Z) \subseteq \mathbb{R}$, we can aggregate restrictions only across (z_D, c) such that $c \in C(z_D; \theta, Z)$. To illustrate, if we aggregate by taking the maximum across (z_D, c) such that $c \in C(z_D; \theta, Z)$, then the induced test statistic and the uncertainty control

[20] Recall from the description after model (1) that the function w_n generates the exogenous dyadic covariates from individual characteristics, i.e., $Z_{ij} = w_n(Z_i, Z_j)$. And "$\tilde{z}_D \in z_D$" means that the discretized version $\tilde{z}_D$ of the vector of individual-leve covariates $\tilde{z}$ according to Definition 1 belongs to the cell z_D.

[21] Outside the interval $C(z_D; \theta, Z)$, neither $\hat{p}_L^{(n)}$ nor $\hat{p}_U^{(n)}$ varies with c: below it 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 $\delta_{ij}(\theta) \leq c$, so $\hat{p}_L^{(n)} = \overline{Y}(z_D)$ and $\hat{p}_U^{(n)} = 1$.

[22] If $\mathcal{X}$ is unbounded but sign-restricted, the same formula can still help trim away unnecessary thresholds.

statistic will be

$$\widehat{Q}_n^{\text{con}}(\theta) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c \in C(z_D;\theta,Z)} \max\Big\{\hat{p}_L^{(n)}(z_D, c; \theta) - F_\varepsilon(c),\ F_\varepsilon(c) - \hat{p}_U^{(n)}(z_D, c; \theta),\ 0\Big\},$$
$$T_\theta^{\text{con}}(R^{(b)}) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c \in C(z_D;\theta,Z)} \big|\hat{F}_n^{(b)}(c; z_D) - F_\varepsilon(c)\big|,$$

which correspond to the aggregation T that keeps the coordinates with $c \in C(z_D; \theta, Z)$ and discards the rest before taking the maximum; multiplying each coordinate by a (θ, Z)-measurable indicator is nondecreasing, so Theorem 4 applies. Then, let $\hat{c}_{n,1-\alpha}^{\text{con}}(Z; \theta)$ be the $(1-\alpha)$ sample quantile of $\big(T_\theta^{\text{con}}(R^{(b)})\big)_{b=1}^B$, and the validity of confidence sets constructed by test inversion is then guaranteed by Theorem 4. That said, constrained threshold sets need not be combined with the maximum aggregation rule: it can be combined with the improvements below as well.

(ii) Studentization.

Dividing each inequality by its null standard deviation equalizes scale across cells of unequal size, so that in finite sample a 20-dyad cell cannot drastically influence the critical value for a 2,000-dyad one. Define

$$\sigma(z_D, c) := \sqrt{\frac{F_\varepsilon(c)\,(1 - F_\varepsilon(c))}{\hat{m}(z_D)}}, \quad \mathcal{S} := \Big\{(z_D, c) : F_\varepsilon(c)\,(1 - F_\varepsilon(c)) \geq \frac{1}{\hat{m}(z_D)}\Big\},$$

where $\sigma(z_D, c)$ is the standard deviation of the cell frequency $\hat{F}_n(c; z_D)$, the average of $\hat{m}(z_D)$ i.i.d. Bernoulli indicators $\mathbf{1}\{\varepsilon_{ij} \leq c\}$ with success probability $F_\varepsilon(c)$. The set $\mathcal{S}$ selects the subset of (z_D, c) whose null standard deviation is at least above a threshold driven by the granularity of the effective sample size $\hat{m}(z_D)$ in the cell, ruling out tail thresholds c (very close to 0 or 1) that are too noisy to "estimate" given the $\hat{m}(z_D)$ data points. The test

To illustrate, for scalar X_{ij} with $\mathcal{X} = [0, \infty)$, we have

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

statistic and uncertainty control statistic can then be defined as

$$\widehat{Q}_n^{\mathrm{stu}}(\theta) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c:\,(z_D,c)\in\mathcal{S}} \frac{\max\left\{\hat{p}_L^{(n)}(z_D,c;\theta) - F_\varepsilon(c),\ F_\varepsilon(c) - \hat{p}_U^{(n)}(z_D,c;\theta),\ 0\right\}}{\sigma(z_D,c)},$$

$$T_\theta^{\mathrm{stu}}(R^{(b)}) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c:\,(z_D,c)\in\mathcal{S}} \frac{\left|\hat{F}_n^{(b)}(c;z_D) - F_\varepsilon(c)\right|}{\sigma(z_D,c)},$$

which are again defined by a single nondecreasing aggregation rule T: divide each (z_D, c) by $\sigma(z_D, c)$, and only sup over $(z_D, c) \in \mathcal{S}$.

(iii) Berk–Jones Weights.

Studentization reweights (z_D, c) through division by the standard deviation of the cell frequency, approximately equalizing the scales of variations across (z_D, c). An alternative way to equalize uncertainty scales across (z_D, c) is to use the Berk–Jones score (Berk and Jones, 1979), which we describe below.

Specifically, fix a cell z_D with $\hat{m}(z_D)$ dyads and a threshold c. Under the null the number of dyads in the cell whose shock falls below c is

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

by Assumption 3, so the cell frequency is $\hat{F}_n(c; z_D) = N(z_D, c)/\hat{m}(z_D)$ with expectation $F_\varepsilon(c)$. Large-deviation probabilities of the form $\mathbb{P}\big(\hat{F}_n(c; z_D) \ge p\big)$ can be approximated by an exponential form

$$\mathbb{P}\big(\hat{F}_n(c;z_D) \ge p\big) \ =\ \exp\Big\{ -\hat{m}(z_D)\,\mathrm{KL}\big(p \,\big\|\, F_\varepsilon(c)\big) + o\big(\hat{m}(z_D)\big)\Big\} \qquad \text{for } p > F_\varepsilon(c),$$

and symmetrically for $p < F_\varepsilon(c)$, where

$$\mathrm{KL}(p\|u) := p\log\frac{p}{u} + (1-p)\log\frac{1-p}{1-u}$$

is the Bernoulli Kullback–Leibler divergence.[23] Weighting the inequality at (z_D, c) by

[23] Both directions are straightforward. For the upper bound, write $N \sim \mathrm{Binomial}(m, u)$ and apply Markov's inequality to $e^{\lambda N}$: for any $\lambda > 0$,

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

and the bracket is maximized at $e^{\lambda} = \frac{p(1-u)}{u(1-p)}$, where it equals exactly $\mathrm{KL}(p\|u)$, giving $\mathbb{P}(N \ge mp) \le e^{-m\mathrm{KL}(p\|u)}$ for every m and every $p > u$. The method of types supplies the matching lower bound $\mathbb{P}(N \ge$

$\hat{m}(z_D)\,\mathrm{KL}\big(\,\cdot\,\big\|\,F_\varepsilon(c)\big)$ therefore reports it on a *tail-probability* scale, while the tail is exactly where we have the most salient concern about finite-sample uncertainty. In particular, in the tails where $F_\varepsilon(c)\,(1-F_\varepsilon(c))$ is near zero the studentized ratio is unstable, while the KL divergence grows more stably, and hence there is no need to "throw away" tails based on the set $\mathcal{S}$ as constructed in the studentization improvement above in (ii).[24]

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 $[F_\varepsilon(c), 1]$, so weighting each side of the sandwich only where that side is violated yields an aggregation that is nondecreasing in the violation magnitude V,[25] and therefore Theorem 4 applies. Taking the maximum over the constrained threshold sets of (i),

$$\widehat{Q}_n^{\mathrm{BJ}}(\theta) = \max_{z_D \in \hat{\mathcal{Z}}_D} \sup_{c \in C(z_D;\theta,Z)} \hat{m}(z_D) \max\Big\{ \mathrm{KL}\big(\hat{p}_L^{(n)}\,\big\|\,F_\varepsilon(c)\big)\mathbf{1}\big\{\hat{p}_L^{(n)} > F_\varepsilon(c)\big\}, \\ \mathrm{KL}\big(\hat{p}_U^{(n)}\,\big\|\,F_\varepsilon(c)\big)\mathbf{1}\big\{\hat{p}_U^{(n)} < F_\varepsilon(c)\big\}\Big\},$$

with the arguments $(z_D, c; \theta)$ of $\hat{p}_L^{(n)}$ and $\hat{p}_U^{(n)}$ suppressed, and the matching control statistic[26] is

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

with $\hat{c}_{n,1-\alpha}^{\mathrm{BJ}}(Z;\theta)$ similarly defined based on simulations of $T_\theta^{\mathrm{BJ}}(R^{(b)})$. This is the constrained-threshold Berk–Jones statistic used in Sections 4 and 5.

$mp) \geq (m+1)^{-1} e^{-m\mathrm{KL}(p\|u)}$ at integer mp, so the exponent is exact. No normal approximation enters at any point.

[24] A concrete illustration. Take a cell of size m and let $u := F_\varepsilon(c) \downarrow 0$ with exactly one dyad below c, so that $p = 1/m$. The studentized deviation is

$$\frac{p-u}{\sqrt{u(1-u)/m}} \;\sim\; \frac{1}{\sqrt{mu}},$$

which diverges at a power rate as $u \downarrow 0$. In comparison, the Berk–Jones weight

$$m\,\mathrm{KL}\big(\tfrac{1}{m}\big\|u\big) = \log\frac{1}{mu} + (m-1)\log\frac{1-1/m}{1-u} \;\sim\; \log\frac{1}{mu},$$

diverges only logarithmically.

[25] Theorem 4 requires T to be nondecreasing in each coordinate of V, and those coordinates are the nonnegative violations of the form $V^{\pm}$. When $\hat{p}_L^{(n)} > F_\varepsilon(c)$ we may write $\hat{p}_L^{(n)} = F_\varepsilon(c) + V^+$, so the assigned weight 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 weight 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^-.

[26] At most one of the two indicators can equal one, since $\hat{p}_L^{(n)} \leq \hat{p}_U^{(n)}$. The same holds on the oracle side, where $\hat{F}_n^{(b)}(c; z_D)$ lies on one side of $F_\varepsilon(c)$ and the two-sided deviation therefore collapses to the single score $\hat{m}(z_D)\,\mathrm{KL}\big(\hat{F}_n^{(b)}(c; z_D)\,\big\|\,F_\varepsilon(c)\big)$.

Remark 7 (Improved Confidence Sets in the Semiparametric Case). The same core idea of this subsection can also be used to improve the semiparametric procedure of Section 3.1, 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

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

At θ_0 the sandwich (11) gives $V \leq R$ coordinate by coordinate, and the statistics (15) and (16) are the special case $T = \max$. Theorem 4 therefore goes through verbatim for any nondecreasing $T_{Z,\theta}$, with the critical value simulated from $R^{(b)}(z_D, z'_D, u) := \left[G_{z_D}^{(b)}(u) - G_{z'_D}^{(b)}(u)\right]_+$ exactly as in Section 3.1.

A key difference from the parametric case, however, is that $T_{Z,\theta}$ cannot be weighted by each c. The semiparametric test exploits the probability integral transform $\hat{F}_n(c; z_D) = G_{z_D}\big(F_\varepsilon(c)\big)$, with simulation run on the transformed latent uniform variable and a supremum over $c \in \mathbb{R}$ replaced by $\sup_{u \in [0,1]}$, which no longer involves c. A weighting scheme that depends on c would interfere with the probability integral transform argument and invalidate the simulation-based test procedure. That said, weights can still depend on Z, in particular the effective sample sizes $\hat{m}(z_D)$ and $\hat{m}(z'_D)$ so that we can down-weigh restrictions coming from cells with very small $\hat{m}(z_D)$ and $\hat{m}(z'_D)$.

Remark 8 (Overidentification and Misspecification Test). Even though the model parameter is in general set identified only, the identifying restrictions still contain *overidentifying restrictions* that can falsify the model specification. In finite sample, the confidence sets of this section control false *exclusion* of the truth under the model. As a result, we may conduct a misspecification test by rejecting correct specification (of the model with all the assumptions) when the $(1 - \alpha)$ confidence set constructed above turns out *empty.*

Corollary 1 (Misspecification Test by Emptiness)**.** *Fix a candidate set* Θ_{grid} *and a significance level* α*. Consider the null hypothesis that the data are generated by model* (1) *under Assumptions 1–4, along with any assumption on* F_ε*, for* some *parameter value in* Θ_{grid}*. Then, for any of the confidence sets of this section, under the null,*

$$\mathbb{P}(\widehat{\mathcal{C}}_n = \varnothing \mid Z) \leq \mathbb{P}(\overline{\theta}_0 \notin \widehat{\mathcal{C}}_n \mid Z) \leq \alpha$$

4 Simulation Study

This section numerically evaluates the performance of the confidence sets proposed in Section 3, across different network sizes from 100 to 10,000.

Table 1: Summary statistics about the simulated networks

n	density mean	density range	mean X	sd X	mean max X
100	.2019	[.1416, .3925]	.0531	.0839	.3237
200	.2002	[.1543, .3379]	.0470	.0705	.2941
400	.2029	[.1617, .2827]	.0458	.0664	.2776
800	.2001	[.1758, .2337]	.0428	.0592	.2487
1600	.1998	[.1826, .2233]	.0419	.0562	.2250
3200	.1994	[.1892, .2152]	.0414	.0548	.2091
6400	.1996	[.1928, .2119]	.0413	.0543	.1980
10000	.1991	[.1931, .2052]	.0410	.0537	.1899

Note: $X = \overline{\text{CF}}$. Based on $K = 100$ replications per size. "mean max X" is the mean across replications of each network's maximum, not the pooled maximum.

4.1 Design

We consider the following model specification

$$Y_{ij} = \mathbf{1}\Big\{\beta_0 \, |z_i - z_j| + \gamma_0 \, \overline{\text{CF}}_{ij}(Y) - b_0 \;\geq\; \varepsilon_{ij}\Big\}, \qquad \beta_0 = -1, \quad \gamma_0 = 16, \tag{22}$$

with ε_{ij} drawn from i.i.d. Logistic$(0, 1)$, z_i drawn from i.i.d. uniform on 21 equally spaced grid on $[-10, 10]$, and $\overline{\text{CF}}_{ij}(Y) := \text{CF}_{ij}(Y)/(n-2)$ 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 $n \in \{100, 200, 400, 800, 1600, 3200, 6400, 10000\}$ with $K = 100$ independent repetitions at every size. We calibrate b_0 via pilot simulation runs so that the outcome networks at $n = 400$ have densities around 0.2, and fix b_0 thereafter.

The choice of the strategic coefficient $\gamma_0 = 16$ seems "large", but its magnitude should be interpreted together with the scale of $X_{ij} := \text{CF}_{ij}(Y)/(n-2)$, which is normalized by $1/(n-2)$. Table 1 reports summary statistics about the equilibrium networks, showing that the endogenous covariate X are generally small in scale with mean around 0.05 and sd around 0.06. 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 $\gamma_0 \cdot \text{mean}(X) \approx 0.75$ logit units at the mean and $\gamma_0 \cdot \text{sd}(X) \approx 1.13$ logit units of variations, against a standard logistic shock ε_{ij}.

We also note that, 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. The focus on supermodular structure enables us to compute large equilibrium networks with $n = 10,000$.

Throughout the rest of this section, we compute various *confidence sets* as described in Section 3 and their performance across the $K = 100$ repetitions. In the parametric case, the standard logistic CDF already fixes the location and scale, so every structural parameter in (β_0, γ_0, b_0) is treated as unknown, and the test (inversion) is joint over the whole vector. We work with a discrete grid on the parameter space: $b_0 \in [-6, 6]$ with step size 0.5, $\beta \in [-3, 1]$ with step size 0.25, and $\gamma \in [-8, 60]$ with step size 0.25. The test statistic is the Berk–Jones statistics in Section 3.3, with the thresholds c fixed on 361 points spanning $[-36, 36]$. The conditional critical value is simulated accordingly with $B = 999$ at confidence level $1 - \alpha = 0.95$. In the semiparametric case, we normalize the scale of the homophily effect parameter to unity, $|\beta_0| = 1$, with the location normalization either imposed via the intercept term b_0 or the symmetry-about-0 assumption.

4.2 Parametric Confidence Sets

Each of the 100 networks at a size is one replication in which the researcher observes that network *alone* and computes a confidence set for the full parameter triple. Table 2 reports statistics of the confidence sets across all three coordinates. We discuss the main findings as follows.

Our inference procedure *is clearly valid and informative* (though conservative as expected), in a single network setting even at the smallest size $n = 100$, and median width of the CS projects shrink at every tested n. The true parameter is covered by the CS in every of the $K = 100$ replications across all n.[27] That said, the CS are nontrivial subsets of the parameter grid that contract with n and deliver sign-revealing information.

Noticeably, the sign of the strategic coefficient γ, an important question of interest in the strategic network formation literature, is certified in 75% of replications at $n = 100$, in 96% at $n = 200$, and in all 100 replications from $n = 400$ onward, while the median γ projection width falls from about 47 (69% of the candidate grid) to about 2 (3.3% of the grid) at $n = 10,000$. Again, each computed confidence set is based on a single network, with no assumption nor information on network equilibrium ever used in the inference procedure. We are unaware of prior inference procedures in the literature that deliver *valid sign-revealing*

[27] This is anticipated since our tests are built on realization-wise inequalities that heuristically ensures the "worst-case" coverage guarantee regardless of network dependence structure and equilibrium selection mechanism.

Table 2: Parametric confidence sets

n	γ sign	γ width	vac.	γ edge L / U	β sign	β width	b_0 width	cover
100	.75	46.750	.07	.14 / .31	.24	2.000	6.500	1.00
200	.96	35.000	.01	.03 / .06	.80	1.250	5.000	1.00
400	**1.00**	22.625	.00	.00 / .02	.98	0.750	3.000	1.00
800	**1.00**	16.250	.00	.00 / .00	**1.00**	0.500	2.000	1.00
1600	**1.00**	10.750	.00	.00 / .00	**1.00**	0.250	1.500	1.00
3200	**1.00**	3.875	.00	.00 / .00	**1.00**	0.000	0.000	1.00
6400	**1.00**	2.750	.00	.00 / .00	**1.00**	0.000	0.000	1.00
10000	**1.00**	2.250	.00	.00 / .00	**1.00**	0.000	0.000	1.00

Note: based on $K = 100$ replications, at 95% nominal level. "signs" is the share of replications whose projection is nonempty with a strictly positive minimum (for γ) or strictly negative maximum (for β). "width" is the median length of the CS projection across replications, and a width of "0.000" means a single grid point at the true value. "vac." is the share accepting all grid points, and "edge" reports contact with the endpoints -8 and 60. "cover" is the share of replications where the CS covers the true parameter vector (jointly).

confidence sets in this environment.

The confidence set also delivers quite tight bounds on the homophily effect parameter β_0 (as well as the intercept b_0). Homophily (negative β_0) is certified in 98% of replications by $n = 400$ and in all of them from $n = 800$. The projection widths for β_0 and b_0 both shrink with n, and from $n = 3200$ onwards the widths around the true values effectively *shrink under the grid resolution* ("0.000"). This confirms the intuition that identification and inference on parameters for the exogenous covariates tend to be "easier". What is particularly noteworthy in our exercise is the finding that this intuition appears still true even in presence of the "hard-to-learn" strategic coefficient γ_0.

4.3 Semiparametric Confidence Sets

Now, in Table 3, we report results on the semiparametric versions of our inference procedure, based on the conditional simulation critical values described in Section 3.1 as well as the symmetry-based one described in Appendix B. We also include corresponding results about the parametric case from the last subsection for the ease of comparison.

These confidence sets are computed on the same simulated networks with the same γ grid as the parametric ones in Section 4.2. A key operational difference lies in the different location and scale normalization between the semiparametric case and the parametric case. The parametric CS are joint over the three-dimensional parameters (b_0, β, γ) jointly, with normalization imposed through the location and scale of the standard logistic distribution.

Table 3: Semiparametric confidence sets: γ projections

n	parametric			symmetric			semiparametric		
	sign	width	vac	sign	width	vac	sign	width	vac
100	.75	46.750	.07	.31	67.250	.45	.28	66.875	.47
200	.96	35.000	.01	.25	66.000	.38	.17	67.375	.45
400	**1.00**	22.625	.00	.61	58.000	.07	.25	59.750	.12
800	**1.00**	16.250	.00	.95	53.625	.00	.38	56.750	.00
1600	**1.00**	10.750	.00	**1.00**	26.500	.00	.73	27.125	.00
3200	**1.00**	3.875	.00	**1.00**	19.750	.00	.95	21.500	.00
6400	**1.00**	2.750	.00	**1.00**	10.750	.00	**1.00**	17.250	.00
10000	**1.00**	2.250	.00	**1.00**	6.500	.00	**1.00**	12.750	.00

Note: "sign": sign rate, "width": median width in the $[-8, 60]$ grid, and "vac": the share of replications accepting the entire grid (vacuous). Vacuity refers to the γ-projection only, not to the joint set. **Joint coverage rates (of truth) are** 100/100 **across all entries** and are thus not reported.

The symmetry-based semiparametric CS (Appendix B) encodes a location normalization (symmetry around 0) but no scale normalization. Consequently, we fix the scale of the homophily effect parameter $|\beta| = 1$ and the symmetry-based CS is joint over an intercept b_0, the strategic coefficient γ_0, as well as the sign of β_0. Lastly, the fully semiparametric CS (Section 3.1) leave both location and scale normalization open: we thus fix $b_0 = 0$ (effectively dropping the constant term) as location normalization and set $|\beta| = 1$ as scale normalization.

The semiparametric confidence sets, while less tight than the parametric ones as anticipated, remain nontrivial and informative. Given the generality and complexity of the strategic network formation problem under consideration, it is *not a priori clear* whether any inference method can deliver nontrivial bounds, especially on the strategic coefficient γ_0. It is thus surprising that our semiparametric confidence sets turn out to be informative on the sign of γ_0, especially given how *few assumptions* were used in the semiparametric inference procedure.

It should be acknowledged that the semiparametric confidence sets still materially get censored by the grid bounds, especially at small n. Upper-endpoint contact is about 80% at $n = 400$ and 27-45% at $n = 800$, so the "width" entries at $n = 400$ and $n = 800$ are contaminated with grid truncation and thus should be interpreted with care. The results at $n = 100$ and $n = 200$ should be read as sign evidence alone. That said, for larger networks (starting $n = 1600$), the semiparametric confidence sets deliver two-sided bounds in the grid interior for the majority of replications.

The "cleanest" metric to focus on is the *sign rate*, the share of replications that certify the

Table 4: Multidimensional exogenous covariates: γ projections

dim-Z	n	parametric			symmetric			semiparametric		
		sign	width	vac	sign	width	vac	sign	width	vac
2	100	.74	34	.00	.00	68	.99	.00	68	1.00
	200	.99	30	.00	.00	68	.51	.00	68	1.00
	400	**1.00**	26	.00	.03	52	.00	.00	68	.70
	800	**1.00**	24	.00	.43	40	.00	.07	54	.00
	1600	**1.00**	24	.00	**1.00**	31	.00	.98	36	.00
	3200	**1.00**	24	.00	**1.00**	26	.00	**1.00**	28	.00
	10000	**1.00**	24	.00	**1.00**	24	.00	**1.00**	26	.00
3	100	.55	38	.00	-					
	200	.93	30	.00	.00	68	1.00	.00	68	1.00
	400	**1.00**	26	.00	.10	68	.63	.00	68	1.00
	800	**1.00**	22	.00	**1.00**	48	.00	.00	68	.97
	1600	**1.00**	20	.00	**1.00**	38	.00	.93	58	.00
	3200	**1.00**	20	.00	**1.00**	30	.00	**1.00**	42	.00
	10000	**1.00**	20	.00	**1.00**	24	.00	**1.00**	28	.00

Note: "sign": sign rate, "width": median width in the $[-8, 60]$ grid, and "vac": the share of replications accepting the entire grid (vacuous). Vacuity refers to the γ-projection only, not to the joint set. "-" indicates too few effective sample sizes in the dicretized cells for the semiparametric procedure to be meaningfully run. **Joint coverage rates (of truth) are** 100/100 **across all entries** and are thus not reported.

(correct) sign of γ_0, and this measure also shows a clear ordering of the confidence set performance with respect to the strength of the imposed assumptions. The strongest parametric assumption delivers the 100% sign rate starting at $n = 400$, the semiparametric symmetry-based one reaches 100% at $n = 1600$, while the purely semiparametric one (without any assumption on F_ε) requires $n = 6400$ to do so.

4.4 Multidimensional Exogenous Covariates

The design so far has one exogenous covariate $|z_i - z_j|$. We now repeat the exercise with two and three exogenous covariates, with true coefficient vectors $\beta_0 = (-1, +1)$ and $\beta_0 = (-1, +1, -.5)$. Each margin of Z_i is independently uniform on the same 21 points in $[-10, 10]$, and the scale normalization is imposed on the first coordinate $|\beta_1| = 1$. The γ grid spans the same $[-8, 60]$, with a larger step size to save computation. The results for the parametric and the two semiparametric confidence sets are reported in Table 4.

We note first that, under the multidimensional Z design, the effective sample sizes in the "discretized cells" z_D often tend to be too small at $n \leq 400$, especially for the 3-dimensional

design. For example, there are 21^3 support points in the 3-dimensional Z space, so at $n = 100$ the parametric CS is very wide (even after cell grouping) and the semiparametric procedure is effectively not meaningful at all, which is thus labeled by "-" in Table 4. The situation improves gradually as with n, but for $n \leq 400$, even the parametric CS remains wide, while the semiparametric versions are mostly vacuous. That said, this is an anticipated issue as the dimension of Z grows.

With the caveat above acknowledged, the results still continue to demonstrate the informativeness of our inference procedure, especially in its ability to certify the sign of γ_0. In addition, the performance ordering with respect to the strength of assumptions on F_ε remains clearly shown even under multidimensional Z. The parametric CS reaches sign rates $\geq 90\%$ at $n = 200$ in both 2-dim-Z and 3-dim-Z settings, while the semiparametric versions eventually reach $\geq 90\%$ rates after $n \geq 1600$.

5 Empirical Application

We apply the parametric confidence sets of Section 3.2 to two empirical data sets: contact (and friendship) among high-school students, and friendship among Twitch streamers. In each network Y_{ij} records the observed link, X_{ij} is a common-friend statistic computed from that same network, and Z_{ij} collects observed dyadic covariates to be explained below. We do not specify which stable network is chosen when the model admits more than one.

5.1 High-School Contact and Friendship Networks

The Thiers13 data contain 327 students in nine classes and four class types (Mastrandrea et al., 2015).[28] 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 to which only about 130 of the 327 students responded (and hence the friendship network should only be interpreted with this caveat). The exogenous covariates are an indicator for different classes and an indicator for different class types, so $Z_{ij} = (1, D_{ij}, T_{ij})$, and the endogenous covariate is the raw common-friend count in the same network. Selected summary statistics are reporeted in Table 5.

Again, we use the Berk–Jones statistic with constrained thresholding in Section 3.3.

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

Table 5: High-school network: summary statistics

network	links	density	mean degree	CF mean (sd)	$\hat{\mathbb{P}}(\text{CF} \geq 1)$
contact, $\underline{\kappa} = 1$	5,818	.1092	35.58	4.334 (6.636)	.7320
contact, $\underline{\kappa} = 2$	4,031	.0756	24.65	2.108 (4.616)	.4091
contact, $\underline{\kappa} = 3$	3,337	.0626	20.41	1.464 (3.591)	.3211
friendship	406	.0076	2.48	0.053 (0.434)	.0247

Note: all four networks have 327 nodes and 53,301 dyads. Common friends in raw integer count over all unordered dyads, including zeros.

Table 6: High-school network: parametric confidence-set projections

network	b_D	b_T	γ
contact, $\underline{\kappa} = 1$	$[-4.0, 3.0]$	$[-4.0, 2.0]$	$[0.03, 0.35]$
contact, $\underline{\kappa} = 2$	$[-4.0, 1.0]$	$[-3.0, 1.0]$	$[0.05, 0.37]$
contact, $\underline{\kappa} = 3$	$[-4.5, 0.5]$	$[-3.5, 0.0]$	$[0.07, 0.39]$
friendship	$[-4.5, 2.5]$	$[-5.0, 3.0]$	$[0.11, 2.24]$

Note: joint nominal confidence level 0.95; each column reports projections of the same joint confidence set. The intercept is jointly tested but its projection is omitted.

Since the raw numbers of common friends CF are already small in scale and in the empirical applications there is just one given sample size n, for convenience we simply use the raw CF count but work with a small-scale grid for γ with step size 0.01 (noticing that the results can be equivalently translated into $CF/(n-2)$ with γ correspondingly scaled up). The remaining coefficient are based on grid with step size 0.5, and the joint grid is chosen so that every displayed projection below lies strictly inside the grid interior.

Table 6 reports the parametric 95%-level confidence-set projection result under the assumption that ε_{ij} is standard logistic.

All four common-friend projections are positive. 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 0.05 to 0.37. At the sample mean of about 2.1 common friends, the endogenous term contributes roughly 0.1 to 0.8 index units, relative to standard logistic error scale. The $\underline{\kappa} = 1$ and $\underline{\kappa} = 3$ rows give nearly the same range, so the positive sign of γ does not depend on a single contact threshold.

The friendship CS projection is also positive but less precise, which is consistent with the fact that "friendship" is constructed from a survey where only 130 of the 327 students responded at all. Correspondingly there is very low variation in the "friendship" data, where

Table 7: Twitch gamer networks: summary statistics

language	nodes	links	density	CF mean (sd)	$\hat{\mathbb{P}}(\text{CF} \geq 1)$
German (DE)	9,498	153,138	.00340	0.863 (2.508)	.3721
English (EN)	7,126	35,324	.00139	0.082 (0.412)	.0631
Spanish (ES)	4,648	59,382	.00550	0.660 (2.141)	.2772
French (FR)	6,549	112,666	.00525	1.093 (2.947)	.3894
Portuguese (PT)	1,912	31,299	.01713	2.175 (5.064)	.5132
Russian (RU)	4,385	37,304	.00388	0.458 (1.516)	.2325

Note: common friends in raw integer count over all unordered dyads, including zeros.

97.5% of dyads have no common friend and 193 of the 327 students are isolated. Again, this means that the results based on the friendship network should be taken with this substantive caveat in mind, and regarded as a supplemental robustness check using this data set.

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).[29] We treat each network as a separate inference problem with language-specific parameters, and each confidence sets is computed on each network separately. See Table 7 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. We construct

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

so that V_{ij} measures the popularity gap, S_{ij} captures the popularity level of the pair, and M_{ij} the (centered) mature-account count. The structural index in language-ℓ network is then

$$\delta_{ij,\ell} = b_{0\ell} + b_{V\ell} V_{ij} + b_{S\ell} S_{ij} + b_{M\ell} M_{ij} + \gamma_\ell \, \mathbf{1}\{\text{CF}_{ij} \geq 1\},$$

with $Z_{ij} = (1, V_{ij}, S_{ij}, M_{ij})$ and the endogenous regressor $X_{ij} = \mathbf{1}\{\text{CF}_{ij} \geq 1\}$ as an indicator of whether the pair has at least one shared neighbor. We work with a full fixed threshold (c) grid from -48 to 48 in steps of 0.25. The parameter grid uses step sizes of .05 on all parameters. Again the joint parameter grid is chosen so that every displayed projection below lies strictly in the grid interior.

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

Table 8: Twitch gamer networks: right tails of CF distribution

language	$\hat{\mathbb{P}}(\mathrm{CF}=0)$	q_{90}	q_{95}	q_{99}	$q_{99.9}$	max CF
German (DE)	.628	2	4	9	24	1,432
English (ENGB)	.937	0	1	2	4	134
Spanish (ES)	.723	2	3	8	24	436
French (FR)	.611	3	5	11	26	1,095
Portuguese (PTBR)	.487	6	9	20	54	449
Russian (RU)	.768	1	2	6	14	562

Note: quantiles of raw CF counts over all unordered dyads, including zeros.

Table 9: Twitch gamer network: parametric confidence-set projections

language	b_V (views gap)	b_S (views level)	b_M (mature pair)	γ
German (DE)	$[0.05, 0.60]$	$[0.20, 0.70]$	$[-0.80, 0.30]$	$[1.45, 7.25]$
English (EN)	$[0.05, 0.50]$	$[0.30, 0.65]$	$[-0.35, 0.35]$	$[0.65, 8.45]$
Spanish (ES)	$[-0.05, 0.85]$	$[0.10, 0.90]$	$[-1.30, 1.10]$	$[1.05, 7.45]$
French (FR)	$[-0.20, 0.85]$	$[0.05, 0.90]$	$[-0.95, 1.00]$	$[0.95, 7.20]$
Portuguese (PT)	$[-0.40, 1.30]$	$[-0.15, 1.10]$	$[-0.70, 0.60]$	$[0.90, 6.95]$
Russian (RU)	$[0.15, 0.95]$	$[0.20, 0.95]$	$[-0.95, 0.95]$	$[0.55, 7.20]$

Note: the six language rows are separate confidence sets at level 0.95 (joint over all parameters), with no parameter homogeneity imposed across networks. The intercept is jointly tested in every row but its uninterpretable projection is omitted.

The reason why we use the binary $X_{ij} = \mathbf{1}\{\mathrm{CF}_{ij} \geq 1\}$ is to control the highly skewed and heterogeneous distribution of CF_{ij} within each language network and across the six language networks. Table 8 documents how extreme that skewness is. In the German network, for instance, 62.8% of dyads have no common friend at all and the 99th percentile is 9, yet the maximum CF_{ij} is 1,432. A raw count would therefore let a handful of neighborhoods dominate the identifying variation in X_{ij}, while it is equally problematic to use logs to reduce skewness given the large share of zeros in the CF_{ij}. All the six networks are very skewed, but the magnitudes of the skewness vary substantially across the six. As a result, the indicator $\mathbf{1}\{\mathrm{CF}_{ij} \geq 1\}$ seems as natural "common friends" type statistic with clear interpretation and retains stability within and across the six networks.

Table 9 reports the parametric confidence-set projections under the assumption that ε_{ij} is standard logistic. Every language produces a strictly positive CS projection for γ: the lower endpoints range from 0.55 in Russian to 1.45 in German. German has the largest lower endpoint, which under the logistic normalization implies more than a 4.2-fold multiplier of the latent threshold odds associated with the first common friend. The upper endpoints are less informative in these sparse networks with a binary endogenous regressor, though they are

comparable in raw magnitudes (6.95-8.45) across the six languages, indicating the stability of our inference procedure across these very different networks. We thus view the robustly positive lower endpoints and the stable upper endpoints as an empirical demonstration of the informativeness and robustness of our inference procedure with respect to the strategic coefficient γ.

References

Andrews, D. W. and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81, 609–666.

Aradillas-López, A. (2020): "The econometrics of static games," *Annual Review of Economics*, 12, 135–165.

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

Badev, A. (2021): "Nash equilibria on (un)stable networks," *Econometrica*, 89, 1179–1206.

Bajari, P., H. Hong, and S. P. Ryan (2010): "Identification and estimation of a discrete game of complete information," *Econometrica*, 78, 1529–1568.

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.

Besag, J. and P. Clifford (1989): "Generalized Monte Carlo significance tests," *Biometrika*, 76, 633–642.

Chandrasekhar, A. G. and M. O. Jackson (2025): "A network formation model based on subgraphs," *The Review of Economic Studies*, 92, 3741–3787.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: Estimation and inference," *Econometrica*, 81, 667–737.

Comola, M. and A. Dekel (2026): "Estimating network externalities in undirected link formation games," *Journal of Applied Econometrics*, published online, DOI 10.1002/jae.70079; volume/pages not yet assigned.

Davezies, L., X. D'Haultfœuille, and Y. Guyonvarch (2021): "Empirical process results for exchangeable arrays," *The Annals of Statistics*, 49, 845–862.

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.

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.

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, 73–80.

Epstein, L. G., H. Kaido, and K. Seo (2016): "Robust confidence regions for incomplete models," *Econometrica*, 84, 1799–1838.

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 preprint 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.

——— (2020): "Network data," in *Handbook of Econometrics*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, Amsterdam: North-Holland, vol. 7A, 111–218.

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, pages per publisher chapter listing — re-verify at proof stage.

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. Watts (2002): "The evolution of social and economic networks," *Journal of economic theory*, 106, 265–295.

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.

Kaido, H. and Y. Zhang (2025): "Universal Inference for Incomplete Discrete Choice Models," ArXiv preprint arXiv:2501.17973.

Kasy, M., E. Linos, and S. Mobasseri (2026): "Causal inference for social network formation," *arXiv preprint arXiv:2604.17952*.

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.

Leung, M. P. and H. R. Moon (2026): "Normal approximation in large network models," *Review of Economic Studies*, forthcoming.

Li, L. and M. Henry (2026): "Finite Sample Inference in Incomplete Models," ArXiv preprint arXiv:2204.00473 (v4, June 2026).

Li, M., Z. Shi, and Y. Zheng (2026): "Bagging the Network," ArXiv preprint arXiv:2410.23852 (v3, May 2026).

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.

——— (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

Marshall, J. (2026): "The Econometrics of Utility Transferability in Dyadic Network Formation Models," *arXiv preprint arXiv:2603.25641*.

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. (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 (rev. May 2022).

Puelz, D., G. Basse, A. Feller, and P. Toulis (2022): "A graph-theoretic approach to randomization tests of causal effects under general interference," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84, 174–204.

Ridder, G. and S. Sheng (2025): "Two-Step Estimation of a Strategic Network Formation Model with Clustering," ArXiv preprint 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 preprint arXiv:2404.12581.

Yanchenko, E., J. P. Williams, and R. Martin (2026): "Universal Inference for Model Selection on Networks," ArXiv preprint arXiv:2606.30981.

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

A Proofs

Proof of Lemma 1. Fix a dyad type z_D and define, for $c \in \mathbb{R}$,

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

and

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

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

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 $N_n := \binom{n}{2}$ dyads of a bounded kernel of the array $W_{ij} := (Z_i, Z_j, \varepsilon_{ij})$. Under Assumptions 2–4 this array is jointly exchangeable (i.e. the 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,

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

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.,

$$\mathbb{E}[\mathbf{1}\{\varepsilon_{ij} \le c\}\mathbf{1}\{Z_{D,ij} = z_D\}] = F_\varepsilon(c) q(z_D)$$

follows from the independence between ε and Z (Assumption 4). Hence,

$$\hat{F}_n(c; z_D) \to F_\varepsilon(c), \quad \hat{F}^-_n(c; z_D) \to F_\varepsilon(c^-) \tag{23}$$

for each fixed c, almost surely.

Next, we establish uniformity of the convergence of $\hat{F}_n(c; z_D)$ over $c \in \mathbb{R}$ at each z_D. Fix an integer $M \ge 2$ and let $c_j := \inf\{c : F_\varepsilon(c) \ge j/M\}$ for $j = 1, \ldots, M-1$. Each c_j is finite and thus well-defined since $F_\varepsilon(c) \to 0$ as $c \to -\infty$ and $F_\varepsilon(c) \to 1$ as $c \to +\infty$. Clearly,

$$F_\varepsilon(c_j) \;\ge\; j/M \qquad \text{and} \qquad F_\varepsilon(c_j^-) \;\le\; j/M.$$

Let Ω_{M,z_D} be the probability-one event that $\hat{D}_{n,z_D} > 0$ eventually and that the pointwise

convergence results (23) hold at every c_j, and set

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

which converges to 0 on Ω_{M,z_D}. Now, fix any $c \in \mathbb{R}$ and let $j \in \{0, \dots, M-1\}$ be such that $c_j \le c < c_{j+1}$, under the conventions $c_0 := -\infty$, $c_M := +\infty$, $F_\varepsilon(c_0) = \hat{F}_n(c_0; z_D) := 0$, and $F_\varepsilon(c_M^-) = \hat{F}_n^-(c_M; z_D) := 1$. By the monotonicity of $c \mapsto \hat{F}_n(c; z_D)$ and of F_ε, we have

$$\hat{F}_n(c; z_D) \ \le\ \hat{F}_n^-(c_{j+1}; z_D) \ \le\ F_\varepsilon(c_{j+1}^-) + \Delta_{n,z_D} \ \le\ \tfrac{j+1}{M} + \Delta_{n,z_D} \ \le\ F_\varepsilon(c) + \tfrac{1}{M} + \Delta_{n,z_D},$$

where the last inequality follows from $F_\varepsilon(c) \ge F_\varepsilon(c_j) \ge j/M$. Symmetrically

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

where the last inequality uses $F_\varepsilon(c) \le F_\varepsilon(c_{j+1}^-) \le (j+1)/M$. Hence,

$$\sup_{c\in\mathbb{R}} \big|\hat{F}_n(c; z_D) - F_\varepsilon(c)\big| \ \le\ \Delta_{n,z_D} + \tfrac{1}{M} \qquad \text{on } \Omega_{M,z_D}.$$

Finally, we establish the uniformity over $z_D \in \mathcal{Z}_D$, with $\mathcal{Z}_D$ being a finite set with $q(z_D) > 0$ at every z_D by Definition 1. For each fixed $M \ge 2$, let $\overline{\Omega}_M$ now denote the intersection of the probability-one events Ω_{M,z_D} over the finitely many $z_D \in \mathcal{Z}_D$. Then $\overline{\Omega}_M$ still occurs with probability one. Let $\Delta_{n,M} := \max_{z_D\in\mathcal{Z}_D} \Delta_{n,z_D}$. Then still $\Delta_{n,M} \to 0$ on $\overline{\Omega}_M$ as $n \to \infty$, and furthermore we have

$$\max_{z_D\in\mathcal{Z}_D} \sup_{c\in\mathbb{R}} \big|\hat{F}_n(c; z_D) - F_\varepsilon(c)\big| \ \le\ \Delta_{n,M} + \tfrac{1}{M} \text{ on } \overline{\Omega}_M.$$

Since M is an arbitrary integer, $\bigcap_{M\ge 2} \overline{\Omega}_M$ is still a probability-one event, under which

$$\max_{z_D\in\mathcal{Z}_D} \sup_{c\in\mathbb{R}} \big|\hat{F}_n(c; z_D) - F_\varepsilon(c)\big| \ \to 0, \quad \text{as } n \to \infty.$$

□

Proof of Theorem 1. Taking $\limsup_n$ on the left of (11) 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

$$p_U^\infty(z_D, c; \theta_0) = \liminf_n \hat{p}_U^{(n)}(z_D, c; \theta_0) \geq \liminf_n \hat{F}_n(c; z_D) = F_\varepsilon(c)$$

which together imply that

$$p_L^\infty(z_D, c; \theta_0) \leq F_\varepsilon(c) \leq p_U^\infty(z_D, c; \theta_0)$$

for all $c \in \mathbb{R}$ and for all z_D. Then

$$\sup_{z_D} p_L^\infty(z_D, c; \theta_0) \leq F_\varepsilon(c) \leq \inf_{z_D} p_U^\infty(z_D, c; \theta_0),$$

which yields part (i). Part (ii) then 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-network sandwich (11) holds cell by cell at θ_0:

$$\hat{p}_L^{(n)}(z_D, c; \theta_0) \;\leq\; \hat{F}_n(c; z_D) \;\leq\; \hat{p}_U^{(n)}(z_D, c; \theta_0) \qquad \text{for every } z_D \in \hat{\mathcal{Z}}_D$$

Maximizing the left inequality and minimizing the right one over $z_D \in \hat{\mathcal{Z}}_D$ yields

$$\max_{z_D} \hat{p}_L^{(n)}(z_D, c; \theta_0) \;\leq\; \max_{z_D} \hat{F}_n(c; z_D), \qquad \min_{z_D} \hat{p}_U^{(n)}(z_D, c; \theta_0) \;\geq\; \min_{z_D} \hat{F}_n(c; z_D),$$

and hence

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

The right-hand side is at most R_n by the definition (16) and is nonnegative. Hence, taking the supremum over c and applying the positive part function preserves the bound, implying $\widehat{Q}_n(\theta_0) \leq R_n$. □

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 size $\hat{m}(z_D)$ are all fixed. Let $(V_{ij})_{ij}$ be i.i.d. $U(0,1)$ random variables, independent of everything else, and define the *distributional transform*

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

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 $U_{ij} \sim \text{Unif}[0,1]$ and $\varepsilon_{ij} = F_\varepsilon^{-1}(U_{ij})$, where $F_\varepsilon^{-1}(u) := \inf\{x : F_\varepsilon(x) \geq u\}$ is the generalized inverse of F_ε. Using the fact that $F_\varepsilon^{-1}(u) \leq c \iff u \leq F_\varepsilon(c)$, we have $\mathbf{1}\{\varepsilon_{ij} \leq c\} = \mathbf{1}\{U_{ij} \leq F_\varepsilon(c)\}$ and hence

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

recalling that 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 3). Hence, G_{z_D} therefore 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 (16) and writing $\mathcal{R} := F_\varepsilon(\mathbb{R}) \subseteq [0,1]$ for the range of F_ε, we have

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

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 $R_n \leq R^*$ is always true and thus R^* remains a valid device to control size as shown below.

Formally, since $R^* \sim R^{(b)}$, we have

$$\mathbb{P}\big(R^* > q^R_{n,1-\alpha}(Z) \,\big|\, Z\big) \leq \alpha,$$

and hence, using $\widehat{Q}_n(\theta_0) \leq R_n \leq R^*$ from Lemma 2, we have

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

In fact, as discussed in Footnote 14, we can deliver exact finite-B guarantee using the alternative critical value $\hat{q}^R_{n,1-\alpha}(Z)$ instead of $q^R_{n,1-\alpha}(Z)$. Specifically, $\hat{q}^R_{n,1-\alpha}(Z)$ is defined as the $\lceil (1-\alpha)(B+1) \rceil$-th smallest value in the simulated sample $R^{(1)}, \ldots, R^{(B)}$. Since $(R^*, R^{(1)}, \ldots, R^{(B)})$ is conditionally i.i.d. given Z, the rank of R^* in $(R^*, R^{(1)}, \ldots, R^{(B)})$ is uniform on the discrete grid $\{1, \ldots, B+1\}$ up to ties. Hence,

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

where any ties only reduce the left-hand side further. Combining with $\widehat{Q}_n(\theta_0) \leq R_n \leq R^*$ again delivers the conditional coverage guarantee. □

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

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

where

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

Obviously, $R_n \le 2\tilde{R}_n$ and hence

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

As before, $\hat{F}_n(\cdot\,; z_D)$ 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 inequality (Dvoretzky et al., 1956; Massart, 1990), which holds for every distribution function, 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 again yields the coverage guarantee. □

Proof of Lemma 3. Condition on Z. Fix a cell z_D and 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 (11) and the

definition of $R_n^{par}(\theta_{\varepsilon,0})$, for every $z_D \in \hat{\mathcal{Z}}_D$,

$$\begin{aligned}\hat{p}_L^{(n)}(z_D, c; \theta_0) - F_\varepsilon(c) &\leq \hat{F}_n(c; z_D) - F_\varepsilon(c) \leq R_n^{par}(\theta_{\varepsilon,0}), \\ F_\varepsilon(c) - \hat{p}_U^{(n)}(z_D, c; \theta_0) &\leq F_\varepsilon(c) - \hat{F}_n(c; z_D) \leq R_n^{par}(\theta_{\varepsilon,0}).\end{aligned}$$

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

Proof of Theorem 3. Condition on Z. As before, by Assumption 4, 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 $R_n^{par}(\theta_{\varepsilon,0})$ and the simulated $R_n^{par,(b)}(\theta_{\varepsilon,0})$ share the same distribution. Combined with Lemma 3, the rest of the proof proceeds in the same way as in the proof of Theorem 2. □

Proof of Theorem 4. At the true (or true under the null) θ_0, the sandwich (11) gives $V^{\pm}(z_D, c; \overline{\theta}_0) \leq R^{\pm}(z_D, c; \theta_{\varepsilon,0})$ for every coordinate and every realization, so monotonicity of $T_{Z,\overline{\theta}_0}$ in each argument yields $T_{Z,\overline{\theta}_0}\big(V(\overline{\theta}_0)\big) \leq T_{Z,\overline{\theta}_0}\big(R(\theta_{\varepsilon,0})\big)$. Conditional on Z, the realized $T_{Z,\overline{\theta}_0}(R)$ and its simulated counterparts $T_{Z,\overline{\theta}_0}(R^{(1)}), \ldots, T_{Z,\overline{\theta}_0}(R^{(B)})$ are again i.i.d., each distributed as $T_{Z,\overline{\theta}_0}\big(R(\theta_{\varepsilon,0})\big)$. The rest of the proof is exactly the same as that of Theorem 2. □

B Semiparametric Inference Based on Symmetry Restrictions

The semiparametric inference procedure developed in Section 3.1 can also be adapted to incorporate nonparametric shape restrictions on F_ε. In this appendix section we focus on *symmetry* restrictions. For simplicity, we restrict attention here to the case where F_ε is continuous.

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

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

Note that once symmetry is imposed, setting the center to be 0 is without loss of generality as a location normalization.[30] Given this observation, any parametric assumption that restricts F_ε to a known symmetric family of distributions, including many commonly used families such as normal, logistic, Laplace, and uniform, becomes a strict special case of the nonparametric symmetry assumption above.

[30] That said, given the symmetry-about-0 assumption, a constant term needs to be included in Z_{ij} so that θ_0 contains a free intercept parameter.

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

$$\widehat{Q}_n^{\text{sym}}(\theta) := \sup_c \left[\max\left\{\hat{a}(c) + \hat{b}(c),\ \hat{a}(c) + \hat{a}(-c),\ \hat{b}(c) + \hat{b}(-c)\right\} - 1 \right]_+. \tag{24}$$

The first term, corresponding to $\hat{a}(c) + \hat{b}(c) \leq 1$, encodes the semiparametric restriction already used in Section 3.1, 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:

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

All three follow from the sandwich (11) together with the identity $\hat{F}_n(c; z_D) = G_{z_D}\left(F_\varepsilon(c)\right)$ established in the proof of Theorem 2, 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 (16) evaluated at u, so the symmetry rung 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: draw $\hat{m}(z_D)$ i.i.d. uniform RVs in each cell to form $G_{z_D}^{(b)}$, and take maximum over $u \in [0, 1]$. This again can be used to construct level-$(1 - \alpha)$ conditional confidence sets by test inversion.