Tractable Identification of Strategic Network Formation Models with Unobserved Heterogeneity*

Wayne Yuan Gao[†] Ming Li[‡] Zhengyan Xu[§]

March 9, 2026

Abstract

We develop a tractable identification approach for strategic network formation models with both strategic link interdependence and individual unobserved heterogeneity (fixed effects). The key challenge is that endogenous network statistics (e.g. number of common friends) enter the link formation equation, while the mapping from model primitives to equilibrium network structure is generally intractable. Our approach sidesteps this difficulty using a "bounding-by-c" technique that treats endogenous covariates as random variables and exploits monotonicity restrictions to obtain identifying information. We derive a system of identifying restrictions based on subnetwork configurations: tetrad-based restrictions that completely eliminate all individual fixed effects, triad-based restrictions that partially difference out fixed effects, and general weighted cycle-based restrictions, along with point identification results. Preliminary simulations show that our approach can deliver informative bounds on the structural parameters.

Keywords: network formation, strategic interaction, fixed effects, partial identification, endogeneity, tetrad, triad, subnetwork
JEL Classification: C31, C57, D85

*We thank Xiaohong Chen, Bryan Graham, Roger Moon, Elie Tamer, and Rui Wang 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
[§]Department of Economics, University of Pennsylvania. Email: xzy24@sas.upenn.edu

1 Introduction

Network formation is a central problem in network economics. From the econometric perspective, the goal is to identify and estimate structural parameters in a model where agents are heterogeneous and strategically interdependent when making linking decisions. Substantial progress has been made on models with either unobserved heterogeneity or strategic interdependence, but combining both features in a tractable framework has remained an open problem.

This paper develops a tractable identification approach for strategic network formation models with unobserved individual fixed effects. We consider a latent utility framework where the formation of a link between agents i and j depends on observed dyadic characteristics, individual fixed effects that capture unobserved degree heterogeneity (such as sociability or popularity), and idiosyncratic pairwise shocks. The link formation decision is allowed to depend on endogenous network statistics arising from the equilibrium of the network formation game, such as the number of common friends between i and j. This accommodates strategic complementarities and other forms of link interdependence that are relevant in many economic applications.

The main methodological challenge is that the mapping from model primitives to equilibrium network structure is generally intractable. In strategic network formation games, the equilibrium network depends on the entire profile of agent characteristics, fixed effects, and shock realizations through complex equilibrium conditions. This gives rise to at least three difficulties. First, the space of possible network structures is a discrete space with a combinatorially overwhelming number of elements[1], making it very hard to handle both theoretically and computationally. Second, solution concepts for network formation problems (such as pairwise stability) usually do not have uniqueness properties, and sometimes the number of equilibria can be very large (de Paula, 2020). Third, iterative procedures are in general not guaranteed to converge to an equilibrium network even when such an equilibrium does exist (Jackson and Watts, 2002). Consequently, characterizing the equilibrium mapping, let alone inverting it for identification purposes, is typically infeasible, except in very simple special cases.

This intractability has led the existing literature to either (i) focus on models without strategic interdependence, such as in the seminal work by Graham (2017) and follow-up work by Gao (2020) and Gao, Li, and Xu (2023), or (ii) study strategic models under restrictions that eliminate or simplify the role of unobserved heterogeneity (Mele, 2017;

[1] For a standard illustrative example: there are 2^{435} possible (undirected and unweighted) network structures on a set of 30 agents.

de Paula, Richards-Shubik, and Tamer, 2018; Sheng, 2020).

Our approach avoids the need for equilibrium characterization by exploiting monotonicity restrictions to obtain identifying information without knowing the form of the equilibrium mapping. The technique, which we call "bounding by c," treats endogenous covariates as random variables and uses indicator function arguments to derive bounds on structural parameters that hold regardless of the equilibrium realization. This approach builds on and extends ideas from Gao and Wang (2026), who developed similar techniques for dynamic panel data models under a partial stationarity condition.

Based on our key technique, we derive a system of identifying restrictions based on different subnetwork configurations. Our primary restrictions exploit tetrad configurations, i.e., sets of four agents with a particular pattern of links, to difference out all individual fixed effects simultaneously. The resulting bounds depend only on the distribution of idiosyncratic shocks, which can be aggregated into identifying restrictions on the model parameters under a standard independence assumption on the idiosyncratic shocks. We also complement the tetrad restrictions with two additional types of restrictions: (1) "incomplete differencing" restrictions, such as triad-based restrictions, that do not eliminate all fixed effects but nevertheless provide identifying information, and (2) "cyclically differencing" restrictions that involve longer cycles of links than those in a tetrad.

Beyond partial identification, we establish conditions under which the structural parameters are point identified. When the pairwise shocks follow a logistic distribution and the endogenous covariates are bilinear in link outcomes—as in the common-friends statistic—we show that a log-odds ratio of tetrad probabilities, conditioned on diagonal link absences that isolate the endogenous covariates from the tetrad shocks, identifies a linear index of the structural parameters. The fixed effects cancel algebraically through tetrad differencing, while the endogenous covariates decouple through the isolation conditioning; these two mechanisms operate independently. The resulting identification equation leads to a computationally simple conditional logit estimator that accommodates both endogenous network statistics and unobserved individual fixed effects, generalizing the tetrad logit of Graham (2017) to settings with strategic link interdependence.

We report a preliminary simulation exercise illustrating the finite-sample performance of the proposed identifying restrictions. We simulate networks under nested specifications that progressively add fixed effects and an endogenous covariate capturing local network structure, and compute the resulting identified set using the tetrad inequalities. The results show that the restrictions can deliver nontrivial bounds on the strategic-interaction parameter even in the full model with endogenous covariates and fixed effects. A larger-scale Monte Carlo study is under investigation.

Related Literature

This paper relates to several strands of the econometric literature on network formation; see de Paula (2020) for a recent survey. The first strand concerns dyadic models with unobserved heterogeneity but without strategic interdependence. Graham (2017) introduced the tetrad logit estimator, which differences out additive fixed effects by comparing link patterns across tetrads of agents. This approach achieves point identification of homophily parameters under a conditional logit specification. Gao (2020) extended these ideas to nonparametric identification of the homophily effect function, while Gao, Li, and Xu (2023) developed logical differencing techniques for settings with non-transferable utilities where bilateral consent is required for link formation. Our paper builds on these ideas but generalizes the setting by allowing endogenous covariates arising from strategic interaction.

A complementary line of recent work develops general fixed-effects methods applicable to network data. Bonhomme and Dano (2023) extend the functional differencing approach of Bonhomme (2012) from panel data to network settings, deriving moment restrictions on model parameters that hold regardless of the form of heterogeneity and without requiring dense networks; their framework, however, treats the network as exogenous. Bonhomme, Dano, and Graham (2025) characterize all moment conditions for nonlinear panel data models that are robust to both unrestricted feedback and arbitrary heterogeneity, with applications to duration and count models. Dano, Honoré, and Weidner (2025) provide a systematic treatment of conditional likelihood and moment-based identification in binary logit models with general fixed effects, covering both panel and network (dyadic) data; their analysis subsumes the tetrad logit of Graham (2017) as a special case. Our paper differs from this body of work in that we address the additional complication of endogenous network covariates arising from strategic interaction, which requires a distinct identification strategy based on tetrad inequalities rather than conditional likelihood or moment equalities.

The second strand studies strategic network formation under various equilibrium concepts. Mele (2017) analyzes exponential random graph models (ERGMs) as potential games with strategic complementarities, developing simulation-based estimation methods. However, ERGMs are known to suffer from degeneracy problems and computational challenges in large networks. de Paula, Richards-Shubik, and Tamer (2018) and Sheng (2020) study partial identification in strategic network formation models under simultaneity, using subnetwork restrictions to bound the identified set. Sheng (2020) provides particularly elegant results using small subnetwork configurations, but her analysis does not accommodate agent-level fixed effects. Menzel (2026) develops a many-agent asymptotic approximation for pairwise stable networks with anonymous and non-anonymous interaction effects, characterizing the limiting distribution of link intensities through a fixed-point system. His identification of payoff parameters requires parametric specification of the utility function and solution of the equilibrium fixed point, and does not address individual fixed effects of the type we consider here. Our approach addresses the fixed-effects gap by combining the differencing logic from the dyadic literature with techniques that accommodate strategic interdependence, while avoiding the need for equilibrium computation.

On the inference side, Leung (2019) and Leung and Moon (2025) develop laws of large numbers and central limit theorems for network moments in sparse strategic formation models embedded in a latent position space, using a branching-process subcriticality condition to control strategic dependencies. Menzel (2021) provides an alternative central limit theorem based on the exchangeability of agents' potential values, which does not require positional homophily or K-locality and applies to general D-adic network moments. In Section 4, we provide primitive conditions under which the conditional probabilities underlying our identifying restrictions can be consistently estimated from a single large network, drawing on these frameworks.

To our knowledge, this paper is the first in the econometric literature to provide identification results for network formation models that allow for both (i) unobserved individual fixed effects and (ii) endogenous network statistics arising from strategic link formation. Our contributions are as follows. First, we show that tetrad-based configurations can be used to difference out all individual fixed effects even when endogenous covariates are present, yielding bounds on structural parameters that are informative under reasonable support conditions. Second, we develop complementary "incomplete differencing" restrictions that deliver additional identification power by combining the two traditional "difference-out" and "aggregate-out" approaches for the handling of fixed effects in the econometric literature. Third, we demonstrate that the "bounding by c" technique from the panel data literature in Gao and Wang (2026) can be productively adapted to network settings, suggesting broader applicability of the key idea.

Methodologically, our paper builds upon the recent work on semiparametric identification with non-separable unobservables. The "bounding by c" technique we employ shares conceptual foundations with the partial stationarity approach of Gao and Wang (2026) for dynamic panel models, and the "multi-inequality aggregation" also relates to the logical operations underlying Gao, Li, and Xu (2023). The commonality across these papers is the use of arithmetic and logical operations based on monotonicity and inequality conditions. In particular, the "bounding by c" technique in Gao and Wang (2026) and the present paper provides a way to "eliminate" the potentially complicated endogenous variables when establishing identifying restrictions based on homogeneity assumptions on the structural error terms.

Finally, our paper relates to the broader literature on games with incomplete information and heterogeneous agents. Aguirregabiria and Mira (2007) and Bajari, Benkard, and Levin (2007) develop estimation methods for dynamic games that avoid full solution of the game by using conditional choice probability representations. While our setting differs (we study a static network formation game rather than a dynamic game), the spirit of avoiding intractable equilibrium computations through the use of observable implications is similar. That said, the exact techniques we exploit to achieve this in our network formation context are naturally different from theirs.

The remainder of the paper is organized as follows. Section 2 presents the model setup, introduces notation, and states our maintained assumptions. Section 3 derives the main identifying restrictions, beginning with tetrad-based restrictions, then developing triad-based and weighted alternatives, and culminating in a general cycle-based differencing framework. Section 4 provides primitive sufficient conditions under which the high-level identifying assumptions hold, embedding the model in the sparse network framework of Leung (2019). Section 5 establishes conditions for point identification under logistic errors and a bilinear endogenous covariate structure. Section 6 provides simulation results based on tetrad-based identifying restrictions. Section 7 concludes.

2 Model and Assumptions

2.1 Model Setup

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. Throughout, we index distinct unordered pairs as (i, j) with $i < j$; all dyadic quantities (Y_{ij}, Z_{ij}, X_{ij}, ε_{ij}) are understood to be symmetric in their indices.

We consider the following network formation model with both link interdependence and fixed effects, where a link between agents i and j exists if and only if the latent surplus from the link is non-negative:

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

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 (e.g., $Z_{ij} = |Z_i - Z_j|$ for homophily effects);

- 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.

- A_i and A_j are individual-level unobserved fixed effects capturing heterogeneity in sociability, popularity, or other unobserved characteristics (here A_i acts as a threshold, so larger A_i reduces the link probability; equivalently, one may place $-A_i$ on the left-hand side, recovering the usual interpretation where higher sociability promotes link formation);

- ε_{ij} is an idiosyncratic pairwise shock.

The covariate function w_n may depend on the network size n; this generality is essential for accommodating sparse network asymptotics (Section 4). Specifically, when agents are embedded in a latent space with positions Ξ_i and the network is sparse with $O(1)$ expected degree, the sparsity scaling $r_n \to 0$ rescales positions to $\bar{\Xi}_i := r_n^{-1}\Xi_i$, and the covariate function takes the form $w_n(Z_i, Z_j) = \tilde{w}(\bar{\Xi}_i, \bar{\Xi}_j, W_i, W_j)$ for a *fixed* function $\tilde{w}$ that does not depend on n. The n-dependence of w_n thus enters only through the position rescaling, while the structural parameters $\theta_0 = (\beta_0, \gamma_0)$ remain fixed—analogous to the spatial weight matrix W_n in spatial autoregressive models. This structure ensures that the conditional link probability $\mathbb{P}(Y_{ij} = 1 \mid Z_{ij} = z)$ is an $O(1)$ quantity for any fixed z in the support of Z_{ij}, even though the marginal link probability vanishes under sparsity. In many specifications w_n does not actually depend on n, but we allow it for generality.

Remark 1 (Individual-Level versus Dyadic Rescaling)**.** *In the sparse network framework of Leung (2019), the sparsity scaling $r_n = (\kappa/n)^{1/d_\xi}$ enters through the rescaled homophily measure $\delta_{ij} = r_n^{-1}\|\Xi_i - \Xi_j\|$, which is a function of the* dyadic *distance. We adopt an equivalent but notationally simpler formulation in which the rescaling is absorbed into the individual-level characteristics. Specifically, we define the rescaled type*

$$\bar{Z}_i \ := \ (r_n^{-1}\Xi_i,\ W_i)$$

and write the dyadic covariate as $Z_{ij} = \tilde{w}(\bar{Z}_i, \bar{Z}_j)$ for a fixed *function $\tilde{w}$ that does not depend on n. The n-dependence of the model is thereby localized in the deterministic, individual-level transformation $Z_i \mapsto \bar{Z}_i$, while the structural objects—the parameters $\theta_0 = (\beta_0, \gamma_0)$ and the covariate function $\tilde{w}$—remain invariant to the network size.*

The two formulations are mathematically identical: writing $w_n(Z_i, Z_j) := \tilde{w}(\bar{Z}_i, \bar{Z}_j)$ recovers the size-dependent dyadic function used in Leung (2019). In particular, because $\bar{Z}_i$ is *a deterministic function of Z_i and (Z_i, A_i) is i.i.d. across agents (Assumption 1), the augmented type $\tilde{Z}_i := (\bar{Z}_i, A_i)$ remains i.i.d., and all arguments based on exchangeability and the branching-process domination in Leung (2019) carry over without modification. Likewise, the identification results of Theorems 1–2 are unaffected, since they are purely algebraic and hold for each fixed n regardless of how the n-dependence is parametrized.*

We prefer the individual-level formulation for two reasons. First, it cleanly separates the sparsity mechanism (the rescaling $\Xi_i \mapsto r_n^{-1}\Xi_i$, which governs the rate at which distant agents become unable to interact) from the identification mechanism (the function $\tilde{w}$ and the conditioning on ζ_{ijhk}, which are n-invariant). Second, it makes transparent that the conditional link probability $\mathbb{P}(Y_{ij} = 1 \mid Z_{ij} = z)$ is an $O(1)$ quantity for any fixed z in the support of Z_{ij}: the vanishing marginal link probability $\mathbb{P}(Y_{ij} = 1) = O(n^{-1})$ arises entirely from the thinning of the position distribution under rescaling, not from any n-dependence in the structural model.

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

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

so that the link formation equation becomes

$$Y_{ij} = \mathbf{1}\{v_{ij} \leq \delta_{ij}\}. \tag{3}$$

A distinguishing feature of our model is the presence of the endogenous covariates X_{ij} and the fixed effects A_i, A_j. In particular, the endogenous covariates X_{ij} may arise as functions of the realized network. Specifically, we allow:

$$X_{ij} = \phi_{ij}(Y, Z) \tag{4}$$

where ϕ_{ij} is a known function mapping the realized network Y and exogenous characteristics $Z = (Z_1, \ldots, Z_n)$ to the dyadic covariate. For example, a component of X_{ij} might be the number of common friends between i and j:

$$\mathrm{CF}_{ij} := \sum_{k \neq i,j} Y_{ik} Y_{jk} \tag{5}$$

which captures transitivity effects in network formation. Other examples include measures of local clustering, degree statistics, and their rescaled or normalized variants.

Since X_{ij} depends on the entire network Y, and Y is determined by equation (1) simultaneously for all pairs, the model describes a strategic network formation game under transferable utilities with pairwise stability as the solution concept. In general, the game may have multiple equilibria for a given realization of primitives (Z, A, ε), where $A = (A_1, \ldots, A_n)$ and $\varepsilon = (\varepsilon_{ij})_{i<j}$ collect the fixed effects and idiosyncratic shocks respectively. We represent the realized network abstractly as:

$$Y = g(Z, A, \varepsilon; \theta_0) \tag{6}$$

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.

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. As will be shown below, our identification approach does not require characterization of g: we extract identifying information using monotonicity restrictions that hold regardless of the complexity of the equilibrium correspondence and any equilibrium selection mechanisms.

2.2 Assumptions

We maintain the following assumptions throughout the analysis.

Assumption 1 (Random Sampling)**.** *The pair (Z_i, A_i) is independently and identically distributed across agents $i = 1, \ldots, n$.*

Assumption 1 is standard in the network formation literature. Note that arbitrary dependence between Z_i and A_i within an agent is allowed, as is arbitrary dependence between A_i and ε_{ij}. The latter flexibility is inessential for the tetrad-based restrictions in Section 3, since the tetrad differencing construction eliminates all fixed effects algebraically, but it avoids imposing unnecessary restrictions on the model.

Assumption 2 (Exogeneity of Z_i)**.** *The vector of exogenous characteristics $Z = (Z_1, \ldots, Z_n)$ is independent of the idiosyncratic shocks $\varepsilon = (\varepsilon_{ij})_{i<j}$.*

Assumption 2 is an exogeneity condition requiring that the observables Z are independent of the unobserved pairwise shocks. Note that we do *not* assume that X is independent of ε: since X is a function of the equilibrium network, it is endogenous and generally correlated with the idiosyncratic shocks. The key distinction is between the exogenous covariates Z (which satisfy independence) and the endogenous covariates X (which do not).

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

Assumption 3 is another standard assumption in the network formation literature on the idiosyncrasy of link-level surplus shocks.

Assumption 4 (Consistent Estimation of Tetrad Conditional Probabilities)**.** *Let* $ijhk$ *denote a tetrad of distinct agents and write* $Z_S := (Z_i, Z_j, Z_h, Z_k)$ *for their individual-level exogenous characteristics. Let* E_{ijhk} *be an event in the observable tetrad* σ*-algebra*

$$\sigma_{ijhk} := \sigma\left(Y_{i'j'}, X_{i'j'}, Z_{i'} : i', j' \in \{i, j, h, k\}\right).$$

Then the conditional probability $\mathbb{P}(E_{ijhk} \mid Z_S = z_S)$ *can be consistently estimated from observable data in a single large network.*

Assumption 4 is a high-level regularity condition asserting that conditional probabilities of tetrad events can be consistently estimated from a single large network. The conditioning is on the individual-level characteristics (Z_i, Z_j, Z_h, Z_k), which is consistent with the general subnetwork formulation (Assumption $4'$ below). Since the link formation equation (1) depends on agents i and j only through the dyadic covariate $Z_{ij} = w_n(Z_i, Z_j)$, it is convenient to define the shorthand

$$\zeta_{ijhk} := (Z_{ij}, Z_{hk}, Z_{ik}, Z_{jh}) \tag{7}$$

for the vector of dyadic covariates determined by the tetrad. Because ζ_{ijhk} is a deterministic function of Z_S, any conditional probability $\mathbb{P}(E \mid \zeta_{ijhk} = \zeta)$ is well-defined under the finer conditioning of Assumption 4, and we use the shorthand $\mathbb{E}[\cdot \mid \zeta]$ throughout Section 3 for readability.

This assumption is satisfied when the network exhibits sufficiently weak dependence so that a Law of Large Numbers (LLN) applies to tetrad statistics. We view Assumption 4 as a convenient high-level condition for presenting the identification arguments; Section 4 provides primitive sufficient conditions under which it holds for the class of strategic formation games considered here, drawing on the network limit theory of Leung (2019).

Remark 2 (Population Identification versus Single-Network Estimation)**.** *The interplay between identification and estimation in network models differs from the standard cross-sectional setting, and it is useful to clarify the distinction.*

In classical econometrics with i.i.d. data, the data-generating process (DGP) is fixed, identification asks whether the structural parameters are pinned down by the population distribution, and estimation recovers them from a growing sample drawn from that same distribution.

In our setting, the data consists of a single *network of n agents, and the game-theoretic structure means that the joint distribution of (Y, X, Z) is inherently n-dependent: adding agents changes the strategic environment and hence the equilibrium mapping g in (6). When the covariate construction w_n depends on n (as required for sparse network asymptotics; see Section 4), the observables are defined relative to the current network size.*

Accordingly, it is useful to distinguish two layers. Population-level identification *is a per-n statement: for a game of fixed size n, the bounding-by-c restrictions derived in Section 3 hold whenever the conditional probabilities $p_L(\zeta, c; \theta)$ and $p_U(\zeta, c; \theta)$ are well-defined. This layer is purely algebraic—it relies on monotonicity and the exogeneity of Z (Assumption 2), with no reference to asymptotics, sparsity, or limit theory. In particular, the identified set—which we denote $\Theta_I^{(n)}$ to emphasize its dependence on n (see, e.g., Θ_I^{tetrad} in Theorem 1)—is well-defined for each network size n, and the true parameter satisfies $\theta_0 \in \Theta_I^{(n)}$ for all n, regardless of whether the network is dense or sparse.* Consistent estimation from a single network *is an asymptotic statement: as the network grows ($n \to \infty$), the sample conditional probabilities $\hat{p}(\zeta)$ converge to their population counterparts $p^{(n)}(\zeta)$. This convergence requires a law of large numbers for dependent network statistics, which in turn requires sufficient decay of dependence across the network. Assumption 4 asserts that such convergence holds, and Section 4 provides primitive conditions—in particular, sparsity and subcriticality—under which it can be verified.*

Crucially, the n-dependence of the covariate construction w_n is relevant only at the estimation layer; the identification layer is agnostic about it. Under the regularity conditions of Section 4, the equilibrium "localizes" as $n \to \infty$—the conditional probability of a tetrad event depends only on the $O(1)$-sized strategic neighborhood—so that $p^{(n)}(\zeta)$ stabilizes and the identified sets $\Theta_I^{(n)}$ converge to a well-defined limit.

3 Identifying Restrictions

This section develops the main identifying restrictions for the structural parameters $\theta_0 = (\beta_0, \gamma_0)$. We begin with tetrad-based restrictions that achieve complete elimination of all individual fixed effects, then present additional triad-based restrictions as well as a general class of cycle-based ones.

3.1 Tetrad-Based Restrictions

Our core idea is to combine the "tetrad differencing" technique with the "bounding-by-c" technique together to derive bounds free of potentially complicated endogenous variables and unobserved fixed effects.

Consider a tetrad of distinct agents $ijhk \equiv (i, j, h, k)$. We first explain the tetrad-differencing technique. Specifically, consider the following tetrad event

$$Y_{ij} = Y_{hk} = 1 \text{ and } Y_{ik} = Y_{jh} = 0. \tag{8}$$

which by (3) is equivalent to the event

$$v_{ij} \leq \delta_{ij}, \quad v_{hk} \leq \delta_{hk}, \quad v_{ik} > \delta_{ik}, \quad v_{jh} > \delta_{jh}.$$

which implies the following

$$\Delta v := v_{ij} + v_{hk} - v_{ik} - v_{jh} < \delta_{ij} + \delta_{hk} - \delta_{ik} - \delta_{jh} =: \Delta\delta. \tag{9}$$

We observe that the fixed effects are all differenced out in Δv:

$$\begin{aligned}
\Delta v &= v_{ij} + v_{hk} - v_{ik} - v_{jh} \\
&= (A_i + A_j + \varepsilon_{ij}) + (A_h + A_k + \varepsilon_{hk}) - (A_i + A_k + \varepsilon_{ik}) - (A_j + A_h + \varepsilon_{jh}) \\
&= \varepsilon_{ij} + \varepsilon_{hk} - \varepsilon_{ik} - \varepsilon_{jh} \\
&=: \Delta\varepsilon.
\end{aligned} \tag{10}$$

Since $\Delta v = \Delta\varepsilon$ by (10), the inequality (9) is equivalent to

$$\Delta\varepsilon < \Delta\delta. \tag{11}$$

Now, we combine the above with the "bounding-by-c" technique. Specifically, consider the intersection of the tetrad event (8) with the event $\{\Delta\delta \leq c\}$, i.e.,

$$Y_{ij} = Y_{hk} = 1 \;\text{ and }\; Y_{ik} = Y_{jh} = 0 \;\text{ and }\; \Delta\delta \leq c, \tag{12}$$

which by (11) implies

$$\Delta\varepsilon < \Delta\delta \text{ and } \;\Delta\delta \leq c.$$

The above further implies

$$\Delta\varepsilon \leq c, \tag{13}$$

an inequality on the exogenous $\Delta\varepsilon$ without fixed effects or endogeneity issues.

Let F_Δ denote the CDF of $\Delta\varepsilon$. Since the intersection event (12) implies (13), taking conditional probabilities given $\zeta_{ijhk} = \zeta$ yields

$$\mathbb{E}\left[Y_{ij}Y_{hk}(1-Y_{ik})(1-Y_{jh})\mathbf{1}\{\Delta\delta \le c\} \,|\, \zeta\right] \;\le\; \mathbb{P}\left(\Delta\varepsilon \le c\right) \equiv F_{\Delta}(c) \tag{14}$$

by the independence of ε and Z in Assumption 2.[2]

Similarly, we can obtain another bound by considering the "flipped" event

$$Y_{ij} = Y_{hk} = 0 \;\text{ and }\; Y_{ik} = Y_{jh} = 1 \;\text{ and }\; \Delta\delta > c,$$

which implies $\Delta\varepsilon > \Delta\delta > c$ and hence

$$(1-Y_{ij})(1-Y_{hk})Y_{ik}Y_{jh}\mathbf{1}\{\Delta\delta > c\} \le \mathbf{1}\{\Delta\varepsilon > c\},$$

yielding

$$\mathbb{E}\left[(1-Y_{ij})(1-Y_{hk})Y_{ik}Y_{jh}\mathbf{1}\{\Delta\delta > c\} \,|\, \zeta\right] \;\le\; 1 - F_{\Delta}(c), \tag{15}$$

or equivalently,

$$F_{\Delta}(c) \le 1 - \mathbb{E}\left[(1-Y_{ij})(1-Y_{hk})Y_{ik}Y_{jh}\mathbf{1}\{\Delta\delta > c\} \,|\, \zeta\right]. \tag{16}$$

To state our main results, it is helpful to make explicit the dependence of the index difference on the model parameters and the tetrad $ijhk$. In the following we write the tetrad index difference

$$\Delta_{ijhk}(\theta) := \delta_{ij}(\theta) + \delta_{hk}(\theta) - \delta_{ik}(\theta) - \delta_{jh}(\theta) \tag{17}$$

and two parametrized conditional probabilities (expectations)

$$p_L(\zeta, c; \theta) := \mathbb{E}\left[Y_{ij}Y_{hk}(1-Y_{ik})(1-Y_{jh})\mathbf{1}\{\Delta_{ijhk}(\theta) \le c\} \,|\, \zeta\right], \tag{18}$$

$$p_U(\zeta, c; \theta) := 1 - \mathbb{E}\left[(1-Y_{ij})(1-Y_{hk})Y_{ik}Y_{jh}\mathbf{1}\{\Delta_{ijhk}(\theta) > c\} \,|\, \zeta\right]. \tag{19}$$

We now exploit the above to build identifying restrictions, which have two variants depending whether researchers impose an additional parametric assumption on ε.

We first explain the case where no parametric assumption is made on ε and F_{Δ} is left unknown. In this case, (14) and (16) are not identifying restrictions per se since they involve

[2] Although $\Delta\delta$ involves the endogenous covariates X_{ij} and hence is not $\sigma(Z)$-measurable, the inequality (14) is valid because the bound $\mathbf{1}\{\text{tetrad event}\} \cdot \mathbf{1}\{\Delta\delta \le c\} \le \mathbf{1}\{\Delta\varepsilon \le c\}$ holds *pointwise* (i.e., for every realization of $(Y, X, Z, A, \varepsilon)$), and taking conditional expectations of both sides preserves the inequality. The right-hand side $\mathbb{P}(\Delta\varepsilon \le c \mid \zeta_{ijhk} = \zeta) = F_{\Delta}(c)$ follows from $\varepsilon \perp Z$, since ζ_{ijhk} is a measurable function of $(Z_1, \ldots, Z_n)$.

the unknown $F_\Delta(c)$. However, (14) and (16) together imply

$$p_L(\zeta, c; \theta_0) \ \leq F_\Delta(c) \ \leq p_U(\zeta, c; \theta_0),$$

and then, since the middle term is constant across ζ,

$$\sup_\zeta p_L(\zeta, c; \theta_0) \ \leq F_\Delta(c) \ \leq \inf_\zeta p_U(\zeta, c; \theta_0),$$

which becomes an identifying restriction once we drop the unknown middle term $F_\Delta(c)$. We summarize our result in the following theorem.

Theorem 1 (Tetrad Restrictions with Nonparametric F_Δ)**.** *Under Assumptions 1–4, the true parameter θ_0 belongs to the identified set*

$$\Theta_I^{\text{tetrad}} := \left\{\theta : \ \sup_\zeta p_L(\zeta, c; \theta) \leq \inf_\zeta p_U(\zeta, c; \theta) \text{ for all } c \in \mathbb{R}\right\}, \tag{20}$$

where the supremum and infimum are taken over ζ in the support of ζ_{ijhk}.

Equivalently, define

$$Q^{\text{tetrad}}(\theta) := \sup_{c\in\mathbb{R}} \left\{\sup_\zeta p_L(\zeta, c; \theta) - \inf_\zeta p_U(\zeta, c; \theta)\right\}. \tag{21}$$

Then $\Theta_I^{\text{tetrad}} = \{\theta : Q^{\text{tetrad}}(\theta) \leq 0\}$.

Remark 3 (Outer Region)**.** *The set Θ_I^{tetrad} is an outer region for the true parameter: it contains θ_0 but may also contain parameters θ for which no i.i.d. pairwise shock distribution generates the observed tetrad probabilities. This arises because the construction eliminates $F_\Delta(c)$ by intersecting bounds across ζ, without enforcing the constraint that F_Δ must be the distribution of $\varepsilon_{ij} + \varepsilon_{hk} - \varepsilon_{ik} - \varepsilon_{jh}$ for some i.i.d. pairwise law. Incorporating this convolution structure could tighten the identified set, but we do not pursue this here. Under the parametric logistic specification (Theorem 2), the distribution F_Δ is fully determined, so the parametric identified set does not suffer from this relaxation.*

Next, we consider the case where researchers impose an additional parametric assumption on ε, so that $F_\Delta(c; \theta_\varepsilon)$ is unknown up to a finite-dimensional parameter θ_ε. Practically, if ε is assumed to follow a distribution with location-scale parameters such as the logistic distribution (Graham, 2017) and the normal distribution (Dzemski, 2019), then there is no need to explicitly introduce the parameter θ_ε given the scope of location and scale normalization in model (1). That said, we keep the notation θ_ε for theoretical completeness.

With $F_\Delta(c;\theta_\varepsilon)$ known up to θ_ε, the inequalities (14) and (16) become identifying restrictions themselves, which we summarize in the following theorem.

Theorem 2 (Tetrad Restrictions with Parametric F_Δ)**.** *Suppose Assumptions 1–4 hold and that F_Δ belongs to a known parametric family indexed by $\theta_\varepsilon \in \Theta_\varepsilon$, with CDF $F_\Delta(\,\cdot\,;\theta_\varepsilon)$. Write $\overline{\theta} := (\theta, \theta_\varepsilon)$ for the stacked parameter and $\overline{\theta}_0 := (\theta_0, \theta_{\varepsilon,0})$ for its true value. Then $\overline{\theta}_0$ belongs to the identified set*

$$\overline{\Theta}_I^{\text{tetrad}} := \Big\{\overline{\theta} : \ \sup_\zeta p_L(\zeta, c; \theta) \le F_\Delta(c;\theta_\varepsilon) \le \inf_\zeta p_U(\zeta, c; \theta) \ \textit{for all } c \in \mathbb{R}\Big\}. \tag{22}$$

Equivalently, define the criterion

$$Q^{\text{tetrad}}(\overline{\theta}) := \sup_{c\in\mathbb{R}} \ \max\left\{\sup_\zeta p_L(\zeta, c;\theta) - F_\Delta(c;\theta_\varepsilon), \ F_\Delta(c;\theta_\varepsilon) - \inf_\zeta p_U(\zeta, c;\theta)\right\}. \tag{23}$$

Then $\overline{\Theta}_I^{\text{tetrad}} = \big\{\overline{\theta} : \ Q^{\text{tetrad}}(\overline{\theta}) \le 0\big\}$.

Several features of Theorems 1 and 2 are worth noting. First, the tetrad construction eliminates all individual fixed effects by purely algebraic differencing within a four-agent configuration, so identification does not rely on sufficient-statistic arguments (as in parametric logit) nor on restrictions such as $A_i \perp Z_i$. Second, the "bounding by c" step addresses the endogeneity of equilibrium network statistics by treating the endogenous index components as random variables and conditioning on observable events that imply inequalities in $\Delta\varepsilon$; in particular, we never need to model, estimate, or simulate the conditional distribution of X_{ij}. Third, because the restrictions are derived from monotonicity and equilibrium feasibility alone, they remain valid without computing the equilibrium mapping g and are robust to equilibrium multiplicity and unknown equilibrium selection. We note that the n-dependence of the covariate construction w_n (introduced in Section 2) plays no role in the identification arguments above: the bounding-by-c restrictions hold for each fixed network size n and are purely algebraic. The n-dependence becomes relevant only at the estimation stage, where one must verify that the conditional probabilities p_L and p_U can be consistently estimated; see Remark 2 for a detailed discussion.

3.2 Triad-Based Restrictions

While tetrad restrictions achieve complete elimination of fixed effects, they are not the only source of identifying restrictions. We now develop complementary restrictions based on triads—subnetwork configurations involving three agents. Triad-based restrictions do not fully eliminate all fixed effects, but they have two practical advantages. First, triads are combinatorially more abundant than tetrads ($\binom{n}{3}$ versus $\binom{n}{4}$ configurations), yielding more observations per network. Second, they exploit different aspects of the latent variable structure and can tighten the identified set when combined with tetrad restrictions. We consider several variants below.

Three-Link Triad Restrictions

Consider a triad (i, j, k) and the configuration where agent i is linked to both j and k, but j and k are not linked. Given the event $\{Y_{ij} = Y_{ik} = 1,\, Y_{jk} = 0\}$, the link inequalities imply

$$v_{ij} + v_{ik} - v_{jk} < \delta_{ij} + \delta_{ik} - \delta_{jk}. \tag{24}$$

The left-hand side equals

$$u_{i,jk} := v_{ij} + v_{ik} - v_{jk} = 2A_i + \varepsilon_{ij} + \varepsilon_{ik} - \varepsilon_{jk}, \tag{25}$$

so intersecting with $\{\delta_{ij} + \delta_{ik} - \delta_{jk} \leq c\}$ yields the indicator bound

$$Y_{ij}Y_{ik}(1 - Y_{jk})\mathbf{1}\{\delta_{ij} + \delta_{ik} - \delta_{jk} \leq c\} \leq \mathbf{1}\{u_{i,jk} \leq c\}.$$

Unlike the tetrad case, the fixed effect A_i does not cancel. However, $u_{i,jk} = A_i + \varepsilon_{ij} - \varepsilon_{ik}$ depends only on $(A_i, \varepsilon_{ij}, \varepsilon_{ik})$, which by Assumptions 1–2 are independent of (Z_j, Z_k). Consequently, the distribution of $u_{i,jk}$ does not depend on (Z_j, Z_k) at all, and in particular is invariant to the identity of i. Let G_3 denote this common (unconditional) CDF. Taking expectations conditional on $Z_{jk} = z_{jk}$ and then taking suprema over z_{jk} yields the moment-inequality bounds stated below.

Proposition 1 (Three-Link Triad Bounds)**.** *For any $c \in \mathbb{R}$:*

$$\sup_{z_{jk}} \mathbb{E}\left[Y_{ij}Y_{ik}(1 - Y_{jk})\mathbf{1}\{\delta_{ij} + \delta_{ik} - \delta_{jk} \leq c\} \mid z_{jk}\right] \leq G_3(c) \tag{26}$$

$$G_3(c) \leq 1 - \sup_{z_{jk}} \mathbb{E}\left[(1 - Y_{ij})(1 - Y_{ik})Y_{jk}\mathbf{1}\{\delta_{ij} + \delta_{ik} - \delta_{jk} > c\} \mid z_{jk}\right] \tag{27}$$

One may then incorporate the information in Proposition 1 into the characterization of the identified set following similar manners as in Theorems 1 and 2.

Two-Link Triad Restrictions

An even simpler restriction compares two links originating from a common agent. Specifically, fix i and consider the event that i links to j but not to k. Given the observable event

$\{Y_{ij} = 1,\, Y_{ik} = 0\}$, the link inequalities imply

$$v_{ij} - v_{ik} < \delta_{ij} - \delta_{ik}.$$

Since $v_{ij} - v_{ik} = (A_i + A_j + \varepsilon_{ij}) - (A_i + A_k + \varepsilon_{ik})$, the fixed effect A_i cancels, leaving

$$u_{i,j,k} := v_{ij} - v_{ik} = A_j + \varepsilon_{ij} - A_k - \varepsilon_{ik}. \tag{28}$$

Using the "bounding-by-c" technique and intersecting the above with $\{\delta_{ij} - \delta_{ik} \leq c\}$, we obtain

$$Y_{ij}(1 - Y_{ik})\mathbf{1}\{\delta_{ij} - \delta_{ik} \leq c\} \leq \mathbf{1}\{u_{i,j,k} \leq c\},$$

the right-hand side of which does not depend on Z_i, so its CDF is common across z_i. Let G_2 denote the CDF of $u_{i,j,k}$. Replicating the "flipped" argument, taking expectations conditional on $Z_i = z_i$, and then taking the supremum over z_i, we obtain the following.

Proposition 2 (Two-Link Triad Bounds)**.** *For any $c \in \mathbb{R}$:*

$$\sup_{z_i} \mathbb{E}\left[Y_{ij}(1 - Y_{ik})\mathbf{1}\{\delta_{ij} - \delta_{ik} \leq c\} \mid z_i\right] \leq G_2(c) \tag{29}$$

$$G_2(c) \leq 1 - \sup_{z_i} \mathbb{E}\left[(1 - Y_{ij})Y_{ik}\mathbf{1}\{\delta_{ij} - \delta_{ik} > c\} \mid z_i\right]. \tag{30}$$

where the supremum is over z_i in the support of Z_i.

3.3 Weighted Differencing

The tetrad and triad restrictions developed above use equal weights on links. A natural extension is to consider weighted combinations of link indicators, which can generate additional identifying restrictions by exploiting different linear combinations of the latent inequalities. We illustrate the idea with a single example; the systematic development of weighted and more general differencing schemes is presented in Section 3.4.

Within a tetrad (i, j, k, ℓ), consider the link configuration:

$$Y_{ij}Y_{ik}(1 - Y_{i\ell}) \tag{31}$$

The event $\{Y_{ij} = Y_{ik} = 1,\, Y_{i\ell} = 0\}$ implies $v_{ij} \leq \delta_{ij}$, $v_{ik} \leq \delta_{ik}$, and $v_{i\ell} > \delta_{i\ell}$. Summing these inequalities with weights $(+1, +1, -2)$, we obtain

$$v_{ij} + v_{ik} - 2v_{i\ell} < \delta_{ij} + \delta_{ik} - 2\delta_{i\ell}.$$

Since $v_{ij}+v_{ik}-2v_{i\ell} = (A_j+\varepsilon_{ij})+(A_k+\varepsilon_{ik})-2(A_\ell+\varepsilon_{i\ell})$, by intersecting with $\{\delta_{ij}+\delta_{ik}-2\delta_{i\ell} \leq c\}$ we can derive the bound

$$Y_{ij}Y_{ik}(1-Y_{i\ell})\mathbf{1}\{\delta_{ij}+\delta_{ik}-2\delta_{i\ell} \leq c\} \leq \mathbf{1}\{(A_j+\varepsilon_{ij})+(A_k+\varepsilon_{ik})-2(A_\ell+\varepsilon_{i\ell}) \leq c\}. \quad (32)$$

Let G_{w} denote the CDF of the weighted latent index on the right-hand side,

$$G_{\mathrm{w}}(c) := \mathbb{P}\left((A_j+\varepsilon_{ij})+(A_k+\varepsilon_{ik})-2(A_\ell+\varepsilon_{i\ell}) \leq c\right).$$

Taking expectations conditional on $Z_i = z_i$ and using $\varepsilon \perp Z$ yields a bound whose right-hand side equals $G_{\mathrm{w}}(c)$ and is therefore common across z_i.

Proposition 3 (Weighted Differencing Bound)**.** *Within a tetrad (i,j,k,ℓ), for any $c \in \mathbb{R}$:*

$$\sup_{z_i} \mathbb{E}\left[Y_{ij}Y_{ik}(1-Y_{i\ell})\mathbf{1}\{\delta_{ij}+\delta_{ik}-2\delta_{i\ell} \leq c\} \mid z_i\right] \leq G_{\mathrm{w}}(c). \quad (33)$$

$$G_{\mathrm{w}}(c) \leq 1 - \sup_{z_i} \mathbb{E}\left[(1-Y_{ij})(1-Y_{ik})Y_{i\ell}\mathbf{1}\{\delta_{ij}+\delta_{ik}-2\delta_{i\ell} > c\} \mid z_i\right]. \quad (34)$$

The potential advantage of weighted differencing is that it generates a richer family of moment inequalities, indexed by the weights, which may tighten the identified set beyond what is achievable with the equal-weight tetrad and triad restrictions alone.

3.4 General Weighted Cycle-Based Differencing

It should now be clear that our core identification idea can be generalized further beyond the identifying restrictions developed in Sections 3.1–3.3, which are all instances of a unified principle: assign signed weights to links in a subgraph, and when the weighted sum of link indicators factors into a product of indicators, the "bounding by c" technique yields bounds involving a weighted sum of latent variables. Fixed effects cancel at any agent whenever the sum of weights on links incident to that agent equals zero. We now formalize this observation.

Definition 1 (Weighted Link Configuration)**.** *Let $S = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$ be a subset of distinct agents. A* weighted link configuration *on S consists of a (finite) set of links E_S together with a weight function $\omega : E_S \to \mathbb{Z}$ (or $\mathbb{R}$), where $\omega_e > 0$ indicates a "present" link and $\omega_e < 0$ indicates an "absent" link. The* weighted incidence sum *at agent $i \in S$ is*

$$\sigma_i := \sum_{e \in E_S : i \in e} \omega_e.$$

where $i \in e$ means that i is one of the two agents involved in link e.

Clearly, a weighted link configuration achieves *complete fixed-effect elimination* if $\sigma_i = 0$ for every $i \in S$, i.e., if the weights incident to each agent sum to zero. The identifying restrictions from earlier sections correspond to specific weighted link configurations:

1. **Tetrad (Section 3.1):** The 4-cycle $\{ij, jh, hk, ki\}$ with alternating signs $E^+ = \{ij, hk\}$, $E^- = \{ik, jh\}$. Every agent has signed incidence sum $\sigma = 0$: complete fixed-effect elimination.

2. **Three-link triad (Section 3.2):** The triangle $\{ij, ik, jk\}$ with $E^+ = \{ij, ik\}$, $E^- = \{jk\}$. Node i has $\sigma_i = +2$; nodes j, k have $\sigma_j = \sigma_k = 0$. Partial elimination: A_j and A_k cancel but $2A_i$ remains.

3. **Weighted differencing (Section 3.3):** The star $\{ij, ik, i\ell\}$ with weights $(+1, +1, -2)$, which can be decomposed as $E^+ = \{ij, ik\}$ and $E^- = \{i\ell\}$. Node i has $\sigma_i = 0$; the other nodes have nonzero incidence. Partial elimination: A_i cancels but A_j, A_k, and A_ℓ remain.

4. **Hexad (New Example):** Consider the 6-cycle

$$\{i_1 i_2, i_2 i_3, i_3 i_4, i_4 i_5, i_5 i_6, i_6 i_1\},$$

with alternating signs $E^+ = \{i_1 i_2, i_3 i_4, i_5 i_6\}$ and $E^- = \{i_2 i_3, i_4 i_5, i_6 i_1\}$. Every agent has $\sigma = 0$: complete fixed-effect elimination using six agents.

We are now ready to present an umbrella result for general weighted link configurations. Fix a weighted link configuration (E_S, ω) on S, and let

$$E^+ := \{e \in E_S : \omega_e > 0\}, \qquad E^- := \{e \in E_S : \omega_e < 0\}. \tag{35}$$

Given the event $\{\prod_{e \in E^+} Y_e \prod_{e \in E^-}(1 - Y_e) = 1\}$, each $e \in E^+$ implies $v_e \le \delta_e$ and each $e \in E^-$ implies $v_e > \delta_e$. Multiplying the corresponding inequalities by ω_e and summing yields

$$\sum_{e \in E_S} \omega_e v_e < \sum_{e \in E_S} \omega_e \delta_e.$$

Then

$$\sum_{e \in E_S} \omega_e v_e = \sum_{i \in S} \sigma_i A_i + \sum_{e \in E_S} \omega_e \varepsilon_e =: U_S.$$

Hence, the fixed effect of any agent with $\sigma_i = 0$ is differenced out, while agents with $\sigma_i \neq 0$ contribute residual fixed-effect terms. It then follows from Assumptions 1, 2, and 3 that U_S is independent from any Z_i such that $\sigma_i = 0$. Adding the "bounding-by-c" event $\{\sum_{e\in E_S} \omega_e\delta_e \le c\}$ yields a bound in terms of U_S, which translates to the following.

Proposition 4 (General Weighted Cycle-Based Identifying Restrictions)**.** *Let S be a subset of distinct agents and (E_S, ω) be a weighted link configuration on S. Define the subset of agents whose fixed effects are eliminated under (E_S, ω) by*

$$S_0 := \{i \in S : \ \sigma_i = 0\}. \tag{36}$$

Then for any $c \in \mathbb{R}$ and any realization z_{S_0} of Z_{S_0}, we have

$$\mathbb{E}\left[\prod_{e\in E^+} Y_e \prod_{e\in E^-}(1-Y_e)\mathbf{1}\left\{\sum_{e\in E_S}\omega_e\delta_e \le c\right\}\middle| Z_{S_0} = z_{S_0}\right] \le F_{U_S}(c), \tag{37}$$

and

$$\mathbb{E}\left[\prod_{e\in E^+}(1-Y_e) \prod_{e\in E^-} Y_e\mathbf{1}\left\{\sum_{e\in E_S}\omega_e\delta_e > c\right\}\middle| Z_{S_0} = z_{S_0}\right] \le 1 - F_{U_S}(c), \tag{38}$$

where F_{U_S} is the CDF of U_S (and does not depend on z_{S_0}).

To apply the cycle-based restrictions in Proposition 4 and aggregate them over a class $\mathcal{W}$ that involve subnetwork structures larger than tetrads, we need an analogue of Assumption 4 that guarantees identification of the required subnetwork conditional probabilities.

Assumption 4′ (Consistent Estimation of Subnetwork Conditional Probabilities)**.** *Fix a class $\mathcal{W}$ of admissible weighted link configurations (S, E_S, ω). For any $(S, E_S, \omega) \in \mathcal{W}$, let $\mathcal{E}_S$ be an event measurable with respect to the observable subnetwork σ-algebra*

$$\sigma_S := \sigma\big(Y_{ij}, X_{ij}, Z_i : i, j \in S\big).$$

Suppose that the conditional probability $\mathbb{P}(\mathcal{E}_S \mid Z_S = z_S)$ can be consistently estimated from observable data in a single large network.

Theorem 3 (Identified Set via Aggregation over Admissible Weighted Link Configurations)**.** *Fix a class $\mathcal{W}$ of admissible weighted link configurations (S, E_S, ω). For each $(S, E_S, \omega) \in \mathcal{W}$, let S_0 be as in (36), and define, for any $c \in \mathbb{R}$,*

$$p_L^{(S,\omega)}(z_{S_0}, c; \theta) := \mathbb{E}\left[\prod_{e\in E^+} Y_e \prod_{e\in E^-}(1-Y_e)\mathbf{1}\left\{\sum_{e\in E_S}\omega_e\delta_e(\theta) \le c\right\}\middle| Z_{S_0} = z_{S_0}\right],$$
$$p_U^{(S,\omega)}(z_{S_0}, c; \theta) := 1 - \mathbb{E}\left[\prod_{e\in E^+}(1-Y_e) \prod_{e\in E^-} Y_e\mathbf{1}\left\{\sum_{e\in E_S}\omega_e\delta_e(\theta) > c\right\}\middle| Z_{S_0} = z_{S_0}\right].$$

Then, under Assumptions 1–3 and 4′, the true parameter θ_0 belongs to the identified set

$$\Theta_I^{\mathcal{W}} := \Big\{\theta : \sup_{(S,E,\omega)\in\mathcal{W}} \sup_{c\in\mathbb{R}} \big[\sup_{z_{S_0}} p_L^{(S,\omega)}(z_{S_0}, c; \theta) - \inf_{z_{S_0}} p_U^{(S,\omega)}(z_{S_0}, c; \theta)\big] \le 0\Big\}. \tag{39}$$

Remark 4 (More Restrictions versus Stronger Assumption 4′)**.** *Allowing longer cycles and, more generally, larger subnetworks in $\mathcal{W}$ can tighten the identified set by adding additional conditional moment inequalities. However, the cost is that identification and estimation of the resulting subnetwork conditional probabilities requires a correspondingly stronger version of Assumption 4′. In finite sample, involving longer cycles can be more data-demanding because larger configurations are rarer, and furthermore the estimation can be less well-behaved since larger subnetworks suffer from more salient network dependence issues.*

4 Primitive Conditions for Assumption 4

Define the joint surplus from a link between i and j, viewed as a function of the endogenous covariate value x, as

$$V_{ij}(x) \;:=\; Z_{ij}'\beta_0 + x'\gamma_0 - A_i - A_j - \varepsilon_{ij}. \tag{40}$$

Because the endogenous covariate $X_{ij} = \phi_{ij}(Y, Z)$ depends on the realized network, the value of x at which the surplus is evaluated is itself an equilibrium object. Each potential link (i, j) falls into exactly one of three categories, determined entirely by the *exogenous* primitives (Z, A, ε):

(i) **Robustly present:** $\inf_x V_{ij}(x) > 0$—the surplus is positive for every possible value of the endogenous covariate, so the link forms in every equilibrium.

(ii) **Robustly absent:** $\sup_x V_{ij}(x) \le 0$—the surplus is non-positive for every possible value, so the link is absent in every equilibrium.

(iii) **Non-robust:** $\sup_x V_{ij}(x) > 0$ and $\inf_x V_{ij}(x) \le 0$—the link status depends on the equilibrium value of X_{ij} and hence on the rest of the network.

Definition 2 (Strategic Neighborhood)**.** *The* non-robustness indicator *for the potential link between i and j is*

$$D_{ij} \;:=\; \mathbf{1}\big\{\sup_x V_{ij}(x) > 0\big\} \cdot \mathbf{1}\big\{\inf_x V_{ij}(x) \le 0\big\},$$

where the supremum and infimum are over $x \in [-\bar{x}, \bar{x}]^{d_x}$. Let C_i denote i's connected component in the graph with adjacency matrix $\boldsymbol{D} = (D_{ij})$ (the non-robustness graph*), and* *let* $\mathbf{\Pi}$ *be the graph with adjacency matrix* $\Pi_{ij} = \mathbf{1}\{\inf_x V_{ij}(x) > 0\}$ *(the* robust-link graph*). The* strategic neighborhood *of agent* i *is the set*

$$C_i^+ := \bigcup_{j \in C_i} \mathcal{N}_{\mathbf{\Pi}}(j, 1),$$

i.e., C_i *together with all agents who are robust neighbors of some member of* C_i*.*

The connected components C_i of the non-robustness graph $\boldsymbol{D}$ partition the agent set $\mathcal{N}_n$, but the strategic neighborhoods C_i^+ can overlap: an agent who is a robust neighbor of two distinct non-robustness components belongs to both of their strategic neighborhoods. Nevertheless, strategic neighborhoods localize the equilibrium: the link outcomes within any strategic neighborhood C^+ depend only on the primitives of agents in C^+. To see why, consider a non-robust link (a, b) with $a, b \in C$ for some component C. The endogenous covariate $X_{ab} = \phi_{ab}(Y, Z)$ depends on other link outcomes, but only links incident to a or b contribute. Any such link (a, k) is either robustly absent ($Y_{ak} = 0$, no contribution), robustly present ($k \in \mathcal{N}_{\mathbf{\Pi}}(a, 1) \subset C^+$, value determined by primitives of $\{a, k\} \subset C^+$), or non-robust ($k \in C_a = C$, status determined by the equilibrium within C^+). Hence the equilibrium on C^+ forms a self-contained fixed-point system in the primitives of agents in C^+ (Leung, 2019, Theorem 1). We formalize this via the following assumptions.

Assumption 6 (Local Externalities)**.** *The endogenous covariate function* ϕ_{ij} *in* (4) *depends only on links incident to* i *or* j*: that is,* $X_{ij} = \phi_{ij}(Y, Z)$ *is measurable with respect to* $\sigma\big(\{Y_{ik}, Y_{jk} : k \in \mathcal{N}_n\}, Z\big)$*.*

Assumption 6 is satisfied by common friends $\text{CF}_{ij} = \sum_k Y_{ik} Y_{jk}$, the Jaccard index, degree statistics, and other standard endogenous covariates used in the network formation literature.

Assumption 7 (Bounded Endogenous Covariates)**.** *The endogenous covariate takes values in a known compact set:* $X_{ij} \in [-\bar{x}, \bar{x}]^{d_x}$ *almost surely for some constant* $\bar{x} < \infty$ *that does not depend on* n*.*

Unscaled common-friends counts $\text{CF}_{ij} = \sum_k Y_{ik} Y_{jk}$ can grow with network size in dense networks. In the sparse regime of Assumption 9 below, however, expected degree is $O(1)$, so the expected number of common friends is also $O(1)$. The Jaccard index takes values in $[0, 1]$ by construction, and normalized common friends CF_{ij}/n are bounded as well. For unnormalized count statistics in denser networks, explicit trimming or normalization would be needed to satisfy Assumption 7.

Assumption 8 (Decentralized Selection). *For any $n \in \mathbb{N}$ and any strategic neighborhood C^+, the equilibrium selection mechanism satisfies $Y_{C^+} = \eta_{|C^+|}(\tilde{Z}_{C^+}, \varepsilon_{C^+})$, where $\eta_{|C^+|}(\cdot)$ depends only on the augmented types and shocks of agents in C^+.*

Assumption 8 requires that equilibrium play within each strategic neighborhood is determined locally. As discussed in Leung (2019), this is satisfied by myopic best-response dynamics, which are widely used in the theoretical and econometric literature on network formation.

Assumption 9 (Bounded Strategic Neighborhoods). *The strategic neighborhoods satisfy $\mathbb{E}[|C_i^+|] = O(1)$ as $n \to \infty$.*

Assumption 9 ensures that chains of strategic dependencies do not propagate through the network. It follows from the branching-process domination argument in Theorem 1 of Leung (2019) under sparsity and subcriticality. Intuitively, exploring C_i^+ via breadth-first search is akin to growing a subcritical branching process that almost surely terminates.

Proposition 5 (Packing Argument for Assumption 4). *Suppose Assumptions 1–3, 6–9 hold, and that Z_i has finite support. Then Assumption 4 is satisfied: for every $z_S = (z_i, z_j, z_h, z_k)$ in the support of Z_S and every tetrad event $E_{ijhk} \in \sigma_{ijhk}$, the conditional probability $\mathbb{P}(E_{ijhk} \mid Z_S = z_S)$ can be consistently estimated from a single large network.*

The proof is given in Appendix A.

Remark 5 (From Pointwise to Uniform Convergence). *Proposition 5 establishes pointwise consistency of the conditional probability estimator for fixed conditioning values. The operational criterion $Q^{\text{tetrad}}(\theta)$ involves suprema over ζ and c, and evaluation at varying θ, which requires uniform convergence guarantees not established here. Closing this gap—for instance via empirical process theory for network-dependent data, or by reducing the continuous c-supremum to a finite grid using monotonicity of F_Δ—is an important direction for the formal econometric theory of inference in this setting. For the present paper, we focus on the population-level identification results and treat the simulation evidence as illustrative.*

Remark 6 (Continuous Individual Characteristics). *The finite-support condition in Proposition 5 is imposed on the individual-level characteristics Z_i. When $Z_i = (\Xi_i, W_i)$ is continuously distributed, this can be accommodated by discretizing the individual-level rescaled position $\bar{\Xi}_i$ into finitely many spatial bins, producing a discrete individual-level type $\bar{Z}_i^{\text{disc}}$. The cell estimator then conditions on exact matches in $(\bar{Z}_i^{\text{disc}}, \bar{Z}_j^{\text{disc}}, \bar{Z}_h^{\text{disc}}, \bar{Z}_k^{\text{disc}})$. Any discrete components of (W_i, W_j) are retained as-is, yielding a finite-support individual characteristic and hence a finite-support dyadic covariate Z_{ij}.*

At the population level, the conditional probability $p(z_S) = \mathbb{P}(E_{ijhk} \mid Z_S = z_S)$ is well-defined as a regular conditional expectation at every z_S at which Z_S has positive density; finite support is not required for the object itself to exist. The identifying restrictions of Theorems 1 and 2 likewise hold pointwise for each such z_S (equivalently, for each ζ in the support of ζ_{ijhk}), regardless of whether the support is finite or continuous.

Finite support enters only at the estimation stage: the cell estimator and the packing argument both require exact matches $\{Z_{S,t} = z_S\}$, which have positive probability only when Z_i is discrete. Such discretization is without loss for validity—the identifying restrictions hold for every type-quadruplet—though it is conservative: coarser bins pool heterogeneous covariate values, yielding fewer distinct conditioning values and hence fewer moment inequality restrictions, which may widen the identified set. Finer discretization recovers more identifying power, approaching the continuous-Z_i identified set in the limit.

5 Point Identification

The identified sets in Theorems 1 and 2 are defined by conditional moment inequalities and are therefore generally set-valued. It is natural to ask when these restrictions are sharp enough to yield point identification. In this section we give conditions under which $\theta_0 = (\beta_0, \gamma_0)$ is point identified, exploiting the algebraic structure of the logistic distribution.

When $\gamma_0 \neq 0$, two complications arise. First, the endogenous covariates $X_{ij} = \phi_{ij}(Y, Z)$ create dependence among the four tetrad link outcomes, so the product formula for $\mathbb{P}(\text{tetrad} \mid \zeta, A)$ breaks down. Second, the tetrad index difference $\Delta\delta = \Delta Z'\beta_0 + \Delta X'\gamma_0$ depends on A through the equilibrium, so the conditional-on-A odds ratio varies with A and cannot be factored out when integrating over A. We handle both problems by *conditioning on the realized endogenous covariate values* and restricting attention to *isolated tetrads* in which the endogenous covariates for the four tetrad links do not depend on any tetrad link outcome.

5.1 Setup and Definitions

Fix a tetrad of distinct agents (i, j, h, k). Recall that the four *tetrad links* are $E_t := \{ij, hk, ik, jh\}$; the remaining two links within the quadruplet are the *diagonal links* $\{ih, jk\}$. Write Y_{-E_t} for the collection of all link outcomes outside E_t and $\varepsilon_{E_t} := (\varepsilon_{ij}, \varepsilon_{hk}, \varepsilon_{ik}, \varepsilon_{jh})$ for the tetrad shocks. Define the *tetrad endogenous covariate vector*

$$X_t := (X_{ij}, X_{hk}, X_{ik}, X_{jh}) \tag{41}$$

and the *tetrad exogenous covariate vector* ζ_{ijhk} as in (7).

5.2 Tetrad Isolation

We introduce two structural conditions that together guarantee that the four tetrad links are conditionally independent logistic Bernoulli trials, given the observable covariates.

Assumption 10 (Bilinear Endogenous Covariate)**.** *The endogenous covariate* $X_{ij} = \phi_{ij}(Y, Z)$ *depends on link outcomes* Y *only through cross-products* $Y_{il}Y_{jl}$ *for* $l \neq i, j$*. Formally, there exists a function* ψ *such that*

$$X_{ij} = \psi\big(\{Y_{il}Y_{jl}\}_{l\neq i,j},\ Z\big).$$

Assumption 10 is satisfied by common friends $\text{CF}_{ij} = \sum_{l\neq i,j} Y_{il}Y_{jl}$, normalized common friends CF_{ij}/n, and any weighted triadic-closure measure $\sum_l w_l \cdot Y_{il}Y_{jl}$. It is *not* satisfied by the Jaccard index $\text{CF}_{ij}/(|N(i) \cup N(j)|)$, whose denominator depends on individual degrees (linearly, rather than bilinearly, in Y).

Assumption 11 (Tetrad Exogeneity)**.** *For tetrads in the admissible set* $\mathcal{T}_n$ *(defined below), the non-tetrad link outcomes* Y_{-E_t} *are measurable with respect to* $\sigma(Z, A, \varepsilon_{-E_t})$*, where* $\varepsilon_{-E_t} := \{\varepsilon_{ab} : (a, b) \notin E_t\}$*.*

Assumption 11 requires that the tetrad-specific shocks ε_{E_t} do not affect any link outcome outside the tetrad. Under the strategic-neighborhood framework of Section 4, a sufficient condition is that no agent outside the tetrad belongs to a non-robustness component containing a tetrad agent: formally, $C_a \cap \{i, j, h, k\}^c = \emptyset$ for each $a \in \{i, j, h, k\}$, where C_a is agent a's connected component in the non-robustness graph $\boldsymbol{D}$. This ensures that all non-robust links involving tetrad agents are links *among* the four tetrad agents themselves, so perturbing ε_{E_t} cannot propagate outside the tetrad. Note that this condition does *not* require the four agents' strategic neighborhoods to be mutually disjoint—agents i and j may share robust neighbors, and the link Y_{ij} itself may be non-robust, provided the non-robustness component containing i and j lies entirely within $\{i, j, h, k\}$. Under the bounded-strategic-neighborhood condition (Assumption 9), this isolation condition holds for a positive fraction of quadruplets.

Definition 3 (Admissible Tetrad)**.** *A tetrad* (i, j, h, k) *is* admissible *if the two diagonal links are absent:*

$$Y_{ih} = 0 \quad \textit{and} \quad Y_{jk} = 0.$$

The set of all admissible tetrads satisfying Assumption 11 is denoted $\mathcal{T}_n$*.*

In a sparse network with expected degree $O(1)$, each link exists with probability $O(n^{-1})$, so the diagonal-absence condition holds with probability approaching 1 for any given quadruplet. Consequently, the number of admissible tetrads is $\Theta(n^4)$, and after the greedy packing argument of Proposition 5 one retains $\Theta(n)$ independent tetrads—the same asymptotic order as in the partial-identification setting.

The key property of admissible tetrads is the following lemma, which shows that diagonal absence and the bilinear structure together decouple the tetrad endogenous covariates from the tetrad link outcomes.

Lemma 1 (Tetrad Non-Interaction). *Suppose Assumptions 6 and 10 hold. For any admissible tetrad $(i, j, h, k) \in \mathcal{T}_n$, each endogenous covariate X_e for $e \in E_t$ satisfies*

$$X_e = \phi_e(Y_{-E_t}, Z), \tag{42}$$

i.e., it does not depend on any tetrad link outcome.

The proof is given in Appendix A.

5.3 Point Identification without Fixed Effects

We first present the result in the simpler setting without individual fixed effects ($A_i \equiv 0$ for all i), which isolates the role of the endogenous covariates. Both results in this section require a strengthened version of Assumption 4 that conditions on the realized endogenous covariates and the diagonal link absences.

Assumption 4″ (Augmented Tetrad Estimation). *The conclusion of Assumption 4 continues to hold when the conditioning is augmented to $(Z_S,\ X_t,\ Y_{ih} = 0,\ Y_{jk} = 0)$, where X_t is the tetrad endogenous covariate vector defined in (41).*

Proposition 6 (Point Identification without Fixed Effects). *Suppose Assumptions 1–3, 10–11, and 4″ hold with $A_i \equiv 0$ for all i and $\varepsilon_{ij} \overset{\text{iid}}{\sim} \text{Logistic}(0, 1)$. Define*

$$p_+(\zeta, x_t) := \mathbb{P}\big(Y_{ij}Y_{hk}(1 - Y_{ik})(1 - Y_{jh}) = 1 \bigm| \zeta,\ X_t = x_t,\ Y_{ih} = 0,\ Y_{jk} = 0\big), \tag{43}$$

$$p_-(\zeta, x_t) := \mathbb{P}\big((1 - Y_{ij})(1 - Y_{hk})Y_{ik}Y_{jh} = 1 \bigm| \zeta,\ X_t = x_t,\ Y_{ih} = 0,\ Y_{jk} = 0\big). \tag{44}$$

Then for every (ζ, x_t) in the support,

$$\log \frac{p_+(\zeta, x_t)}{p_-(\zeta, x_t)} \;=\; \Delta Z(\zeta)'\beta_0 + \Delta X(x_t)'\gamma_0, \tag{45}$$

where $\Delta Z := Z_{ij} + Z_{hk} - Z_{ik} - Z_{jh}$ and $\Delta X := X_{ij} + X_{hk} - X_{ik} - X_{jh}$. If the support of $(\Delta Z, \Delta X)$ restricted to admissible tetrads spans $\mathbb{R}^{d_\beta + d_\gamma}$, then $\theta_0 = (\beta_0, \gamma_0)$ is point identified.

The proof is given in Appendix A.

5.4 Point Identification with Fixed Effects

The result extends to the full model with individual fixed effects. The tetrad differencing eliminates all fixed effects *algebraically*—exactly as in the $\gamma_0 = 0$ case—while the endogenous covariates are handled by the isolation conditioning. The two mechanisms do not interfere with each other: fixed effects cancel regardless of the endogenous covariates, and endogenous covariates decouple regardless of the fixed effects.

Theorem 4 (Point Identification with Endogenous Covariates and Fixed Effects)**.** *Suppose Assumptions 1–3, 10–11, and 4″ hold, and* $\varepsilon_{ij} \overset{\text{iid}}{\sim} \text{Logistic}(0,1)$*. Define* $p_+(\zeta, x_t)$ *and* $p_-(\zeta, x_t)$ *as in* (43)–(44)*. Then for every* (ζ, x_t) *in the support,*

$$\log \frac{p_+(\zeta, x_t)}{p_-(\zeta, x_t)} \;=\; \Delta Z(\zeta)'\beta_0 + \Delta X(x_t)'\gamma_0. \tag{46}$$

If the support of $(\Delta Z, \Delta X)$ *restricted to admissible tetrads spans* $\mathbb{R}^{d_\beta + d_\gamma}$*, then* $\theta_0 = (\beta_0, \gamma_0)$ *is point identified.*

The proof is given in Appendix A.

5.5 Constructive Estimator

The proof of Theorem 4 is constructive and yields a computationally simple estimator.

(i) *Select admissible tetrads.* From the observed network, enumerate quadruplets (i, j, h, k) satisfying $Y_{ih} = 0$ and $Y_{jk} = 0$.

(ii) *Classify.* For each admissible tetrad, record whether the tetrad pattern $(Y_{ij}Y_{hk}(1 - Y_{ik})(1 - Y_{jh}) = 1)$ or the flipped pattern $((1 - Y_{ij})(1 - Y_{hk})Y_{ik}Y_{jh} = 1)$ obtains; discard tetrads with neither pattern.

(iii) *Conditional logit.* Among the retained tetrads, run a conditional logit regression of the binary outcome (tetrad $= 1$, flipped $= 0$) on the regressors $(\Delta Z, \Delta X)$:

$$\mathbb{P}(\text{tetrad} \mid \text{tetrad or flipped},\ \zeta,\ x_t) = \frac{\exp\bigl(\Delta Z'\beta + \Delta X'\gamma\bigr)}{1 + \exp\bigl(\Delta Z'\beta + \Delta X'\gamma\bigr)}. \tag{47}$$

The maximum likelihood estimator $\hat\theta = (\hat\beta, \hat\gamma)$ consistently estimates θ_0.

Remark 7 (Comparison with the $\gamma_0 = 0$ Case and Graham (2017))**.** *Theorem 4 generalizes the tetrad logit of Graham (2017) in two directions: it allows endogenous covariates* $(\gamma_0 \neq 0)$*,* *and it uses unconditional (marginal over A) tetrad probabilities rather than probabilities conditional on the sufficient statistic for A. When $\gamma_0 = 0$, there are no endogenous covariates, the diagonal-absence condition is vacuous (since $X_e \equiv 0$ does not depend on any link), and Theorem 4 reduces to the standard tetrad logit with $\log[p_+(\zeta)/p_-(\zeta)] = \Delta Z'\beta_0$. The additional structural restrictions (Assumptions 10–11 and diagonal absence) are the price paid for accommodating $\gamma_0 \neq 0$.*

Remark 8 (Abundance of Admissible Tetrads)**.** *In a sparse network with expected degree $O(1)$, $\mathbb{P}(Y_{ih} = 0) = 1 - O(n^{-1}) \to 1$ and similarly for Y_{jk}. Hence the diagonal-absence condition is satisfied with probability approaching 1 for any given quadruplet, and the number of admissible tetrads is $\Theta(n^4)$. The additional requirement that no outside agent shares a non-robustness component with a tetrad agent (for Assumption 11) is closely related to the condition used in the packing argument of Proposition 5; under Assumption 9, one retains $\Theta(n)$ independent admissible tetrads.*

Remark 9 (Scope of the Bilinear Condition)**.** *Assumption 10 covers the most empirically important endogenous covariate, such as common friends $\mathrm{CF}_{ij} = \sum_{l \neq i,j} Y_{il}Y_{jl}$, and its transformations. For covariates that depend linearly on individual degrees (e.g., the Jaccard index or degree-weighted statistics), the bilinear condition fails because individual degrees such as $|N(i)| = \sum_l Y_{il}$ involve single link outcomes, causing X_{ij} to depend on tetrad links even when the diagonals are absent. Extending point identification to such covariates—e.g., by conditioning on additional link absences or via control-function methods—is an important direction for future work. For the partial identification results, the bounding-by-c technique of Theorems 1–2 remains valid for arbitrary ϕ_{ij} without the bilinear restriction.*

Remark 10 (Rank Condition)**.** *The rank condition requires that $(\Delta Z, \Delta X)$ has $(d_\beta + d_\gamma)$-dimensional variation across admissible tetrads. Since $\Delta X = \mathrm{CF}_{ij} + \mathrm{CF}_{hk} - \mathrm{CF}_{ik} - \mathrm{CF}_{jh}$ (under isolation) depends on the local network structure around the tetrad agents, it provides genuine variation beyond ΔZ as long as the agents have heterogeneous local neighborhoods. When all common-friend counts are zero (e.g., in an extremely sparse network), $\Delta X \equiv 0$ and only β_0 can be identified; the estimator naturally reduces to the standard tetrad logit. In networks with nontrivial clustering, ΔX takes many distinct values, ensuring the rank condition is generically satisfied.*

6 Simulation Based on Tetrad Restrictions

We examine the finite-sample behavior of the tetrad-based partial identification approach through simulation, focusing on how the network size n, unobserved heterogeneity (fixed effects) A, the support size of the discrete exogenous covariate Z, and the endogenous covariate X affect the sharpness of the identified set. For simplicity, we use only the baseline tetrad restrictions (Theorems 1–2), leaving the longer-cycle restrictions and aggregation over $\mathcal{W}$ for future work.

6.1 Model Specification and Data Generating Process

We conduct simulation studies under three nested specifications, which differ by whether individual fixed effects and/or endogenous covariates are present.

Baseline Model. The individual-level exogenous characteristics Z_i is a two-dimensional vector $(Z_{1i}, Z_{2i}) \in [-10, 10]^2$, and the vector of exogenous dyadic covariates is constructed as $Z_{ij} = |Z_i - Z_j|$. The true link formation equation is

$$Y_{ij} = \mathbf{1}\left\{Z_{ij,1} + \gamma_0 Z_{ij,2} + \varepsilon_{ij} \geq 0\right\}. \tag{48}$$

Under all three specifications, the shocks ε_{ij} are i.i.d. following a Logistic$(0, 1)$ distribution.

FE-only Model. We add individual-level unobserved fixed effects $A_i \sim \mathcal{N}(\rho Z_{1i}, \sigma_A^2)$ to the baseline model, where $\rho \in [0, 1]$ captures the correlation between unobserved heterogeneity and observed characteristics. The true link formation equation becomes

$$Y_{ij} = \mathbf{1}\left\{Z_{ij,1} + \gamma_0 Z_{ij,2} + A_i + A_j + \varepsilon_{ij} \geq 0\right\}. \tag{49}$$

Note that this is equivalent to the main model (1) with $\tilde{A}_i := -A_i$ playing the role of the fixed effect (and $-\varepsilon_{ij}$ for the shock, which has the same logistic distribution by symmetry).[3]

Full Model. We further add an endogenous dyadic covariate X_{ij} to the model while reducing the dimension of Z_{ij} to 1, i.e., $Z_i \in \mathcal{Z} = [-10, 10]$ and $Z_{ij} = |Z_i - Z_j|$. We set the coefficient for Z_{ij} to be $\beta_0 = 1$, fixed and known. The endogenous covariate is the Jaccard index of common friends:

$$X_{ij} = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|},$$

[3] Under this reparametrization, $\tilde{A}_i \sim \mathcal{N}(-\rho Z_{1i}, \sigma_A^2)$ and equation (49) becomes $Y_{ij} = \mathbf{1}\{Z_{ij,1} + \gamma_0 Z_{ij,2} + \varepsilon_{ij} \geq \tilde{A}_i + \tilde{A}_j\}$, matching the sign convention in (1). The tetrad restrictions are invariant to this reparametrization since A_i cancels in all cases.

where $N(i) = \{k : Y_{ik} = 1\}$ denotes the neighbor set of i. The true link formation equation becomes

$$Y_{ij} = \mathbf{1}\{Z_{ij} + \gamma_0 X_{ij} + A_i + A_j + \varepsilon_{ij} \geq 0\}. \tag{50}$$

Under this model, the probabilities involved in the identifying restriction (20) no longer have a closed form. Therefore, we need to approximate the probabilities using a realized network.

In order to generate a realized network, we first discretize $\mathcal{Z}$ and generate Z_i from $\mathcal{Z}$ by randomly sampling with equal probabilities. Since X_{ij} depends on the realized network Y, it requires solving for a network consistent with (50) given the shocks and fixed effects. We set $A_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1)$ and generate data by iterating a link-update procedure (starting from a zero Y matrix and recomputing X_{ij} after each update) until convergence. The parameter to be identified is γ_0.

Remark 11 (Simulation Design Choices)**.** *Several design choices merit discussion. The value $\gamma_0 = 4$ was chosen to ensure that the endogenous covariate has a substantial effect on link formation, making the strategic interaction component quantitatively important. Since $X_{ij} \in [0, 1]$, a coefficient of $\gamma_0 = 4$ means that dyads with identical neighborhoods ($X_{ij} = 1$) receive a boost of 4 in the latent index, comparable to the standard deviation of the logistic shock ($\pi/\sqrt{3} \approx 1.81$). Results for moderate values of γ_0 may yield tighter identified sets.*

The best-response iteration starting from the empty network selects one particular equilibrium when multiple equilibria exist. We do not verify uniqueness of the equilibrium; thus our simulation results are conditional on this particular equilibrium selection rule. The identified set derived from our restrictions is valid regardless of equilibrium selection (Theorem 1), but the finite-sample informativeness of the criterion may vary across equilibria.

We use γ to denote the parameter to be identified. Following Theorem 1, we evaluate the (sample analogue of the) tetrad criterion $Q^{\text{tetrad}}(\gamma)$ defined in (21), where the relevant conditional probabilities p_L and p_U are computed (i) in closed form under the baseline model, (ii) by Monte Carlo integration over fixed effects under the FE-only model, and (iii) by sample proportions under the full model. The identified set is the set of values for which the criterion is non-positive, i.e. $Q^{\text{tetrad}}(\gamma) \leq 0$.

Under the full model, we also implement the stronger criterion that enforces both sides of the tetrad restriction under the parametric assumption on the logistic distribution of ε_{ij}, as in Theorem 2; the resulting identified set is denoted by Γ_{strict}.

We evaluate the criterion functions on a finite grid of γ values. The supremum is computed by the GenSA algorithm under the baseline model and the FE-only model, while it is computed by grid-search under the full model.

6.2 Simulation Results

6.2.1 Baseline Model

Figure 1 shows how the value of the main criterion function $Q(\gamma)$ changes with γ between -10 and 10 when the true parameter is $\gamma_0 = 1$. In the figure, Q is plotted after shifting by $+1$, so the identified set $\{\gamma : Q(\gamma) \leq 0\}$ corresponds to the region where the plotted curve attains its minimum value of 1. The identified set for γ is $[1, 5]$.

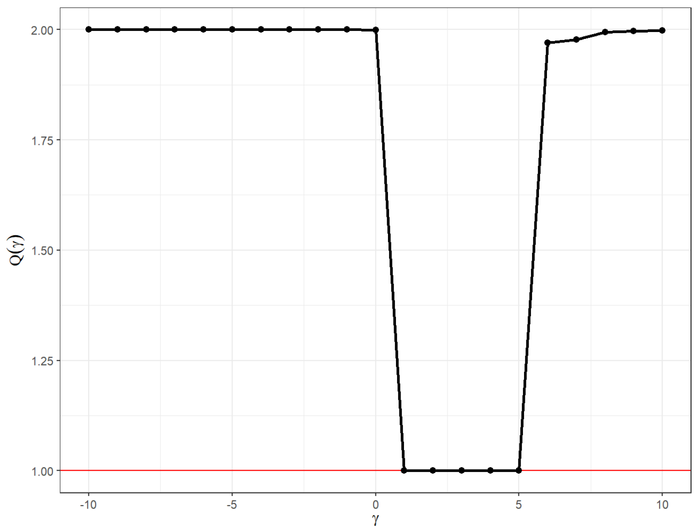

Figure 1: Baseline criterion Q as a function of γ.

6.2.2 FE-only Model

Table 1 reports the identifying results for γ under the FE-only model (49) with fixed-effect designs that vary (i) the dispersion of individual fixed effects and (ii) the strength of correlation between individual fixed effects and observed individual characteristics Z_i.

Table 1: FE-only model: Identified sets for γ under alternative fixed-effect designs.

	(1)	(2)	(3)	(4)
Correlation with Z_{1i}	No	No	Weak	Strong
SD of fixed effects (σ_A)	1	5	1	1
Identified set for γ	$[1, 7]$	$[0, \infty)$	$[1, 6]$	$[1, \infty)$

Notes. The identified set is $\{\gamma : Q(\gamma) \leq 0\}$. In correlated designs, $A_i = \rho Z_{1i} + u_i$ with $u_i \sim \mathcal{N}(0, \sigma_A^2)$; "no", "weak", and "strong" correlation means $\rho = 0$, $\rho = 0.1$, and $\rho = 0.9$, respectively.

Under the FE-only model, introducing unobserved heterogeneity substantially weakens identification of γ relative to the baseline model without fixed effects. In particular, while the baseline model yields a finite identified set $[1, 5]$ when the true value is $\gamma_0 = 1$, the FE-only model produces noticeably wider sets that become sensitive to both the dispersion and endogeneity of fixed effects. When fixed effects are uncorrelated with individual characteristics and have small dispersion, the identified set remains bounded above but expands a little compared to the baseline; increasing heterogeneity makes the upper bound disappear, yielding one-sided identification. Holding dispersion fixed, correlation between the fixed effects and individual characteristics also weakens identification when the correlation is strong. Overall, unobserved heterogeneity, especially when strongly correlated with observables or with large dispersion, reduces the informativeness of the identifying criterion and tends to eliminate finite upper bounds on γ, but the sign of γ can still be identified in all cases.

6.2.3 Full Model

Table 2 reports the identified sets obtained from both the main criterion $Q(\gamma) \leq 0$ and the strict criterion (Theorem 2), using the Jaccard index as the endogenous covariate with $n = 100$ and a single network draw.

Table 2: Full model: Identified sets for γ ($n = 100$, $\gamma_0 = 4$, Jaccard index).

$\lvert\mathcal{Z}\rvert$	$Q(\gamma) \leq 0$	Strict
3	$(-\infty, 28]$	$(-\infty, 24]$
9	$(-\infty, 20]$	$[-26, 18]$
15	$(-\infty, 20]$	$[-5, 19]$
21	$(-\infty, 17]$	$[\mathbf{4}, \mathbf{11}]$

Notes. $(-\infty, \bar{\gamma}]$ indicates a lower endpoint at the search-grid boundary -40. Results are based on a single network draw.

The identified set tightens as the support size $|\mathcal{Z}|$ grows, since more exogenous variation enables more informative tetrad comparisons. At $|\mathcal{Z}| = 21$, the strict identified set is $[4, 11]$, a bounded interval of width 7 containing the true value $\gamma_0 = 4$. This shows that the tetrad restrictions can produce nontrivial bounds on the strategic-interaction parameter even with endogenous covariates and unobserved individual fixed effects.

These results are preliminary: they are based on a single network draw at $n = 100$, and the current design fixes β_0 at its true value and searches only over γ, which does not address the challenges of joint identification over the full parameter vector (β, γ). A more systematic Monte Carlo study—varying network sizes, averaging across repeated draws, searching over the full parameter space, and constructing formal confidence sets for the identified set (e.g., along the lines of Andrews and Shi 2013 and Chernozhukov, Lee, and Rosen 2013)—is under investigation and will be reported in a subsequent version of this paper.

7 Conclusion

This paper develops a tractable identification approach for strategic network formation models with endogenous network statistics and unobserved individual fixed effects. The main idea is a "bounding-by-c" construction applied to subnetwork configurations (tetrads, triads, and more general weighted cycles), which produces moment-inequality restrictions without requiring characterization of the equilibrium mapping. Section 4 gives primitive conditions under which the high-level Assumption 4 holds, embedding the model in the sparse network framework of Leung (2019) and establishing consistency of the tetrad conditional probability estimator via a greedy packing argument.

Several directions for future work are under investigation. First, the central limit theorems of Leung and Moon (2025) and Menzel (2021) can be used to develop formal inference procedures for the identified sets, building on the econometric theory of inference based on conditional moment inequalities (Andrews and Shi, 2013; Chernozhukov, Lee, and Rosen, 2013). Second, we are working on a more systematic set of Monte Carlo simulations in order to provide a fuller picture of the finite-sample performance of our set-identification approach as well as our point identification result (and the associated estimator). Third, we are exploring real-world data sets for an empirical application of our proposed methods.

References

AGUIRREGABIRIA, V. AND P. MIRA (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75, 1–53. 1

ANDREWS, D. W. AND X. SHI (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81, 609–666. 6.2.3, 7

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating dynamic models of imperfect competition," *Econometrica*, 75, 1331–1370. 1

BONHOMME, S. (2012): "Functional Differencing," *Econometrica*, 80, 1337–1385. 1

BONHOMME, S. AND K. DANO (2023): "Functional Differencing in Networks," *Revue Économique*. 1

BONHOMME, S., K. DANO, AND B. S. GRAHAM (2025): "Moment Restrictions for Nonlinear Panel Data Models with Feedback," *arXiv preprint arXiv:2506.12569*. 1

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection bounds: Estimation and inference," *Econometrica*, 81, 667–737. 6.2.3, 7

Dano, K., B. E. Honoré, and M. Weidner (2025): "Binary Choice Logit Models with General Fixed Effects for Panel and Network Data," *arXiv preprint arXiv:2508.11556*. 1

de Paula, Á. (2020): "Econometric models of network formation," *Annual Review of Economics*, 12, 775–799. 1, 1

de Paula, Á., S. Richards-Shubik, and E. Tamer (2018): "Identifying preferences in networks with bounded degree," *Econometrica*, 86, 263–288. 1, 1

Dzemski, A. (2019): "An empirical model of dyadic link formation in a network with unobserved heterogeneity," *Review of Economics and Statistics*, 101, 763–776. 3.1

Gao, W. Y. (2020): "Nonparametric Identification in Index Models of Link Formation," *Journal of Econometrics*, 215, 399–413. 1, 1

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. 1, 1

Gao, W. Y. and R. Wang (2026): "Identification in nonlinear dynamic panel models under partial stationarity," *Journal of Econometrics*, 253, 106185. 1, 1

Graham, B. S. (2017): "An Econometric Model of Network Formation with Degree Heterogeneity," *Econometrica*, 85, 1033–1063. 1, 1, 3.1, 7

Jackson, M. O. and A. Watts (2002): "The evolution of social and economic networks," *Journal of economic theory*, 106, 265–295. 1

Leung, M. P. (2019): "A weak law for moments of pairwise-stable networks," *Journal of Econometrics*, 210, 310–326. 1, 1, 2.2, 4, 4, 4, 7, A.1

Leung, M. P. and H. R. Moon (2025): "Normal approximation in large network models," *arXiv preprint arXiv:1904.11060v5*. 1, 7

Mele, A. (2017): "A Structural Model of Homophily and Clustering in Social Networks," Working paper. 1, 1

Menzel, K. (2021): "Central Limit Theory for Models of Strategic Network Formation," Working paper, New York University. 1, 7

——— (2026): "Strategic network formation with many agents," *Journal of Econometrics*, 253, 106174. 1

Sheng, S. (2020): "A structural econometric analysis of network formation games through subnetworks," *Econometrica*, 88, 1829–1858. 1, 1

A Proofs

A.1 Proof of Proposition 5

Proof. We prove the result using a packing argument. For a tetrad $t = (i, j, h, k)$, define its *dependence set*

$$D_t \;:=\; \bigcup_{a \in \{i,j,h,k\}} C_a^+.$$

This is the set of all agents whose primitives can affect the tetrad outcome E_t. Indeed, by Assumption 8, each link among $\{i, j, h, k\}$ is determined by the equilibrium on the strategic neighborhood containing its endpoints, and all such neighborhoods are contained in D_t. Since the connected components C_a of the non-robustness graph partition $\mathcal{N}_n$, each agent belongs to exactly one component; however, the strategic neighborhoods C_a^+ may overlap. By the union bound and Assumption 9, $\mathbb{E}[|D_t|] \le \sum_{a\in t} \mathbb{E}[|C_a^+|] = O(1)$.

Define a *conflict graph* G_D on the set of all tetrads with $\zeta_t = \zeta$: two tetrads t, t' are adjacent in G_D if and only if $D_t \cap D_{t'} \neq \emptyset$. Since $D_{t'} = \bigcup_{a\in t'} C_a^+$, the tetrads t and t' conflict if and only if at least one agent $a \in t'$ satisfies $C_a^+ \cap D_t \neq \emptyset$. Define the *second-order dependence set*

$$D_t^{(2)} \;:=\; \big\{a \in \mathcal{N}_n : C_a^+ \cap D_t \neq \emptyset\big\}.$$

We now bound the two key quantities:

- *Number of vertices.* Since augmented types $\tilde{Z}_i = (Z_i, A_i)$ are i.i.d. (Assumption 1) and z_S lies in the finite support of Z_S, the number of tetrads with $Z_S = z_S$ is $N = \binom{n}{4}\mathbb{P}(Z_S = z_S) = \Theta(n^4)$.
- *Average degree.* A tetrad t' conflicts with t only if at least one agent of t' lies in $D_t^{(2)}$. Under the subcriticality condition of Leung (2019), $|C_i^+|$ has exponentially decaying tails; iterating the branching-process bound once more shows that $|D_t^{(2)}|$ also has bounded expectation: $\mathbb{E}[|D_t^{(2)}|] = O(1)$. Each agent in $D_t^{(2)}$ can appear in at most $\binom{n-1}{3}$ other tetrads, so the expected degree of t in G_D satisfies $\mathbb{E}[\deg(t)] \le \mathbb{E}[|D_t^{(2)}|]\cdot n^3 = O(n^3)$. Since this bound holds for every tetrad t, the average degree in G_D is $\bar{d} = O(n^3)$.

By the Turán bound, G_D contains an independent set of size at least $N/(\bar{d}+1) \geq \Theta(n^4)/O(n^3) = \Theta(n)$. That is, the greedy algorithm—which sequentially selects any tetrad not conflicting with those previously selected—packs at least $m_n = \Theta(n)$ tetrads $t_1, \ldots, t_{m_n}$ with pairwise disjoint dependence sets. (The greedy selection is deterministic given the realization of (Z, A, ε), so the selected tetrads are random. The key probabilistic properties—independence and identical conditional means—hold for these selected tetrads as shown below.)

Since $D_{t_s} \cap D_{t_{s'}} = \emptyset$ for $s \neq s'$, the primitives determining each E_{t_s} are disjoint subcollections of i.i.d. random variables ($\tilde{Z}_i$ across agents, ε_{ij} across pairs). By Assumption 8, the equilibrium on each strategic neighborhood is determined only by the primitives within it. Moreover, the endogenous covariates X_{ij} for $i, j \in t_s$ depend only on link outcomes incident to i or j (Assumption 6). Any such link (i,k) with $k \notin D_{t_s}$ is robustly absent—since k is not in any strategic neighborhood of a member of t_s, we have $Y_{ik} = 0$—so X_{ij} depends only on primitives within D_{t_s}. Therefore $\{\mathbf{1}\{E_{t_s}\}\}_{s=1}^{m_n}$ are mutually independent.

Moreover, because $\tilde{Z}_i = (Z_i, A_i)$ is i.i.d. across agents (Assumption 1), any two tetrads with $Z_S = z_S$ have the same conditional distribution of outcomes. The identical-mean property holds because each E_{t_s} depends only on primitives within D_{t_s}, and conditional on $Z_{S,t_s} = z_S$, the distribution of these primitives is determined by the i.i.d. structure of $(\tilde{Z}_i, \varepsilon_{ij})$ and does not depend on which tetrad was selected. In particular,

$$\mathbb{E}\big[\mathbf{1}\{E_{t_s}\} \mid Z_{S,t_s} = z_S\big] \;=\; p^{(n)}(z_S) \quad \text{for all } s = 1, \ldots, m_n.$$

Since $\{\mathbf{1}\{E_{t_s}\}\}_{s=1}^{m_n}$ are i.i.d. Bernoulli random variables (conditional on $Z_{S,t_s} = z_S$ for all s) and $m_n \to \infty$, the strong law of large numbers gives

$$\frac{1}{m_n} \sum_{s=1}^{m_n} \mathbf{1}\{E_{t_s}\} \;\xrightarrow{p}\; p^{(n)}(z_S).$$

This establishes that an oracle with access to the packed tetrads could consistently estimate $p^{(n)}(z_S)$. Showing that the natural full-sample average over all observable matched tetrads (a U-statistic) is also consistent requires additional arguments controlling the contribution of dependent tetrad pairs; we leave this formal extension to future work. □

A.2 Proof of Lemma 1

Proof. We verify (42) for each $e \in E_t$. Under Assumption 10, X_e depends on Y only through products $Y_{al}Y_{bl}$ where $e = (a,b)$ and $l \neq a, b$. It therefore suffices to show that for each $e \in E_t$

and each $l \in \{i,j,h,k\} \setminus e$, the product $Y_{al}Y_{bl}$ involves at most one tetrad link—with the other factor being a diagonal link that equals zero.

Covariate X_{ij}: The only within-quadruplet agents are $l \in \{h,k\}$. For $l = h$: $Y_{ih} \cdot Y_{jh} = 0 \cdot Y_{jh} = 0$ (since $Y_{ih} = 0$). For $l = k$: $Y_{ik} \cdot Y_{jk} = Y_{ik} \cdot 0 = 0$ (since $Y_{jk} = 0$).

Covariate X_{hk}: Within-quadruplet agents: $l \in \{i,j\}$. For $l = i$: $Y_{hi} \cdot Y_{ki} = Y_{ih} \cdot Y_{ik} = 0 \cdot Y_{ik} = 0$. For $l = j$: $Y_{hj} \cdot Y_{kj} = Y_{jh} \cdot Y_{jk} = Y_{jh} \cdot 0 = 0$.

Covariate X_{ik}: Within-quadruplet agents: $l \in \{j,h\}$. For $l = j$: $Y_{ij} \cdot Y_{kj} = Y_{ij} \cdot Y_{jk} = Y_{ij} \cdot 0 = 0$. For $l = h$: $Y_{ih} \cdot Y_{kh} = 0 \cdot Y_{hk} = 0$.

Covariate X_{jh}: Within-quadruplet agents: $l \in \{i,k\}$. For $l = i$: $Y_{ji} \cdot Y_{hi} = Y_{ij} \cdot Y_{ih} = Y_{ij} \cdot 0 = 0$. For $l = k$: $Y_{jk} \cdot Y_{hk} = 0 \cdot Y_{hk} = 0$.

For all $l \notin \{i,j,h,k\}$, the products $Y_{al}Y_{bl}$ involve only links to outside agents, hence no tetrad links. □

A.3 Proof of Proposition 6

Proof. Write $\Lambda(x) := (1 + e^{-x})^{-1}$ for the logistic CDF. Fix an admissible tetrad $(i,j,h,k) \in \mathcal{T}_n$. By Lemma 1, each X_e for $e \in E_t$ is a function of (Y_{-E_t}, Z) only. Combined with Assumption 11, (Y_{-E_t}, Z) is measurable with respect to $\sigma(Z, \varepsilon_{-E_t})$. Since $\varepsilon_{E_t} \perp (Z, \varepsilon_{-E_t})$ by Assumptions 2 and 3, we have

$$\varepsilon_{E_t} \ \perp \ (\zeta,\, X_t,\, Y_{ih},\, Y_{jk}).$$

Therefore, conditional on $(\zeta, X_t = x_t, Y_{ih} = 0, Y_{jk} = 0)$, the four tetrad shocks remain i.i.d. Logistic$(0,1)$. Since $A_i \equiv 0$, the threshold for each link is $\delta_e = Z_e'\beta_0 + x_e'\gamma_0$, which is deterministic given (ζ, x_t). The four tetrad links are therefore mutually independent Bernoulli:

$$Y_e \mid (\zeta, x_t, Y_{ih} = 0, Y_{jk} = 0) \ \sim \ \text{Bernoulli}\big(\Lambda(\delta_e)\big), \quad \text{independently for each } e \in E_t.$$

By the product formula for independent Bernoulli trials:

$$\frac{p_+(\zeta, x_t)}{p_-(\zeta, x_t)} = \prod_{e \in \{ij,hk\}} \frac{\Lambda(\delta_e)}{1 - \Lambda(\delta_e)} \cdot \prod_{e \in \{ik,jh\}} \frac{1 - \Lambda(\delta_e)}{\Lambda(\delta_e)}.$$

The logistic odds identity $\Lambda(x)/[1 - \Lambda(x)] = e^x$ then yields

$$\frac{p_+(\zeta, x_t)}{p_-(\zeta, x_t)} = \exp\big[(\delta_{ij} + \delta_{hk}) - (\delta_{ik} + \delta_{jh})\big] = \exp\big(\Delta Z'\beta_0 + \Delta X'\gamma_0\big),$$

giving (45).

The equation $\log[p_+(\zeta, x_t)/p_-(\zeta, x_t)] = (\Delta Z, \Delta X)'(\beta_0, \gamma_0)$ is a system of moment equalities indexed by (ζ, x_t). Under the rank condition, this system has a unique solution. $\square$

A.4 Proof of Theorem 4

Proof. Fix an admissible tetrad $(i, j, h, k) \in \mathcal{T}_n$. The argument of Proposition 6 extends directly. By Lemma 1 and Assumption 11, each X_e $(e \in E_t)$ is measurable with respect to $\sigma(Z, A, \varepsilon_{-E_t})$, and hence $\varepsilon_{E_t} \perp (\zeta, A, X_t, Y_{ih}, Y_{jk})$. Conditional on (Z, A, Y_{-E_t}), the four tetrad links are independent Bernoulli with

$$\mathbb{P}(Y_e = 1 \mid Z, A, Y_{-E_t}) = \Lambda\big(\delta_e - A_{i(e)} - A_{j(e)}\big).$$

By the above and the logistic odds identity:

$$\begin{aligned} \frac{\mathbb{P}(\text{tetrad} \mid Z, A, Y_{-E_t})}{\mathbb{P}(\text{flipped} \mid Z, A, Y_{-E_t})} &= \exp\Big[\Delta\delta - \underbrace{\big((A_i + A_j) + (A_h + A_k) - (A_i + A_k) - (A_j + A_h)\big)}_{=0}\Big] \\ &= \exp\big(\Delta Z'\beta_0 + \Delta X'\gamma_0\big). \end{aligned}$$

The ratio does not depend on A.

Now, since $\exp(\Delta Z'\beta_0 + \Delta X'\gamma_0)$ is constant given (ζ, x_t)—it depends on A neither directly nor through X_t (which is held fixed by the conditioning)—it factors out when taking expectations over (A, Y_{-E_t}) conditional on $(\zeta, x_t, Y_{ih} = 0, Y_{jk} = 0)$:

$$\begin{aligned} p_+(\zeta, x_t) &= \mathbb{E}\big[\mathbb{P}(\text{tetrad} \mid Z, A, Y_{-E_t}) \bigm| \zeta, x_t, Y_{ih} = 0, Y_{jk} = 0\big] \\ &= \exp(\Delta Z'\beta_0 + \Delta X'\gamma_0) \cdot \mathbb{E}\big[\mathbb{P}(\text{flipped} \mid Z, A, Y_{-E_t}) \bigm| \zeta, x_t, Y_{ih} = 0, Y_{jk} = 0\big] \\ &= \exp(\Delta Z'\beta_0 + \Delta X'\gamma_0) \cdot p_-(\zeta, x_t). \end{aligned}$$

Taking logs yields (46).

Finally, the same rank argument as in Proposition 6 delivers the desired result. $\square$