

Homophily and transitivity in dynamic network formation

Kevin Dano, Bryan S. Graham and Yassine Sbai Sassi*

September 12, 2026

*Dano: Department of Economics, Princeton University, Washington Rd, Princeton, NJ 08544, e-mail: kdano@princeton.edu, web: <https://kevindano.github.io>.

Graham: Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880 and National Bureau of Economic Research, e-mail: bgraham@econ.berkeley.edu, web: <http://bryangraham.github.io/econometrics/>.

Sbai Sassi: Department of Economics, New York University 19 West 4th Street, 6th Floor New York, NY 10012, e-mail: ys6799@nyu.edu, web: <https://yassinesbaisassi.github.io/>.

This paper substantially develops and extends ideas initially reported in the second author's working paper "Homophily and transitivity in dynamic network formation". That paper, in turn, expanded upon comments prepared in response to Guido Imbens' *Journal of Business and Economic Statistics Annual Lecture* delivered at the American Economic Association meetings in Chicago (January 7th, 2012). We are grateful to Guido, and then JBES editors Kei Hirano and Jonathan Wright, for providing the initial opportunity to consider the ideas explored in this paper. We thank current JBES editor, Michal Kolesár, for the opportunity to develop this material further. This paper additionally develops previously unreported findings discovered by the first author in the context of dissertation research, a counter-example discovered by the third author, as well as substantial new work done by all three authors collaboratively in preparation for the 2026 JBES lecture in Philadelphia. We are also grateful to Peter Bickel, Paul Goldsmith-Pinkham, Jim Heckman, Bo Honore, Nouredine El Karoui, Michael Leung as well as to participants at the Measuring and Interpreting Inequality (MIE) Inaugural Conference (February 18th, 2012), the CEME Conference on Inference in Nonstandard Problems (June 15th & 16th, 2012), the invited Social Interactions session of the ESEM (August 27 - 31, 2012), the CEMMAP conference on Estimation of Complementarities in Matching and Social Networks (October 5 & 6, 2012), the 9th Invitational Choice Symposium (June 12 - 16, 2013), the Berkeley Statistics Department's NSF reading group on networks, the Third European Meeting on Networks (June 18-19th, 2015), the USC-INET Conference on Networks (November 20-21st, 2015), the 2026 AEA/ES Winter Meetings, Structural Econometrics and Models of Strategic Interactions Conference in Toulouse (May 28th & 29th, 2026) and the UPENN Econometric Seminar for valuable comments. Financial support, for the second author, from NSF Grant SES #1357499 is gratefully acknowledged. Generative AI was used to catch typos, notational inconsistencies and check proofs. We used Gemini 3.1 Pro and GPT-5/6 for these purposes. The authors report there are no competing interests to declare. All the usual disclaimers apply.

Abstract

In social and economic networks linked agents often share connections in common. There are two competing explanations for this phenomenon. First, agents may have a structural taste for transitive links – the returns to linking may be higher if two agents share a common connection. Second, agents may assortatively match on unobserved attributes, a process called homophily. We study parameter identifiability in a simple model of dynamic network formation with both effects. Agents form, maintain, and dissolve links over time to maximize utility. The return to linking may be higher if agents share connections in common. A pair-specific utility component allows for arbitrary homophily on time-invariant agent attributes. We derive conditions under which it is possible to detect the presence of a taste for transitivity in the presence of assortative matching on unobservables. We leave the joint distribution of the initial network and the pair-specific utility component, both high dimensional nuisance parameters, unrestricted. Our identification result is constructive, suggesting an analog estimator, whose finite and (single) large network properties we characterize. We show, via examples, the delicacy of information accumulation in the single (large) network setting.

JEL Codes: C31, C33, C35

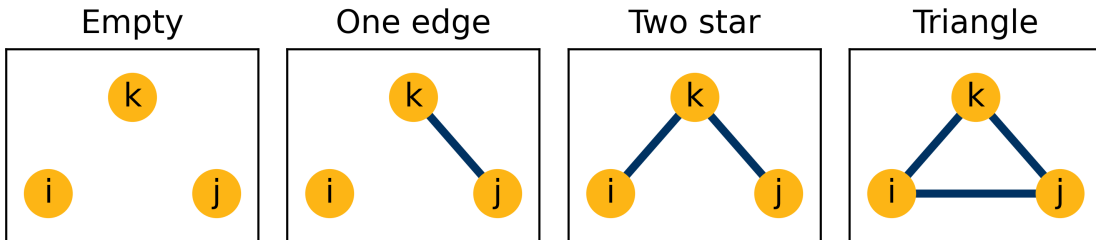
Keywords: Strategic Network Formation, Homophily, Transitivity, Conditional Likelihood, Triads, Assortative Matching, Initial Conditions Problem, Panel Data, Fixed Effects, Incidental Parameters

Let i, j and k index the agents of a *triad*, drawn at random from a simple undirected network. Let $D_{ij} = 1$ if agents i and j in this triad are connected and zero otherwise.¹ The triad ijk must be wired, up to isomorphisms, in one of the four ways shown in Figure 1. The two-star and triangle configurations interest us here. Consider the probability that edge ij is present *conditional* on the presence of edges ik and jk ; that is, the probability that agents i and j are connected given that they share k as a mutual friend:

$$\rho_{CC} = \Pr(D_{ij} = 1 | D_{ik} = 1, D_{jk} = 1). \quad (1)$$

The sample analog of (1) is called the transitivity index or global clustering coefficient in the networks literature (Harary and Kimmel, 1979; Dekker et al., 2019). In real world social and economic networks ρ_{CC} is generally higher, often substantially so, than $\rho_D = \Pr(D_{ij} = 1)$, the unconditional frequency at which agents link (Jackson, 2008). Networks exhibit *transitivity*: agents link more frequently if they have connections in common (“the friend of my friend is also my friend”).

Figure 1: Triad configurations in undirected networks



Measured transitivity reflects two distinct phenomena. First, agents may *prefer* transitive ties. That is, the payoff associated with a link between any two agents may increase with the number of neighbors they share in common. Coleman (1990) argues that closed sub-networks, like triangles, facilitate the formation of social and human capital. Jackson et al. (2012) provide a game-theoretic foundation for transitivity, arguing that common friends, by monitoring transactions between agents, help to sustain cooperation. More mundanely socializing may be easier and more enjoyable when individuals share friends in common.

Second, transitivity may reflect assortative matching on unobserved attributes. Link payoffs may be greater across similar agents, leading to dense ties among them. The tendency for individuals to assortatively match on gender, race, age and other observed characteristics, called *homophily* by network researchers, is widely documented (McPherson et al., 2001).

¹Since links are undirected, $D_{ij} = D_{ji}$ for all pairs ij . Self-ties are also ruled out, so that $D_{ii} = 0$ for all i .

Understanding how to detect a structural taste for transitive ties in the presence of unobserved homophily is the primary subject of this paper. Distinguishing between a taste for transitivity and homophily as drivers of clustering in networks is of considerable scientific interest and policy relevance. When agents prefer transitive links, temporary outside interventions can induce long-run changes in network structure. Whereas, if clustering primarily reflects assortative matching on unobserved agent attributes, such interventions generally do not generate durable changes in network structure (Graham, 2020).

There is a useful analogy between our research problem, both in terms of scientific motivation and technical content, and that of discriminating between state dependence and unobserved heterogeneity in single-agent dynamic binary choice analysis (Cox, 1958; Heckman, 1978, 1981a,b,c; Chamberlain, 1985). The multi-agent aspect of our problem generates new challenges, nevertheless we will utilize intuitions and proof strategies from research in this area (e.g., Honoré and Kyriazidou, 2000).

Section 1, which appears next, introduces a dynamic model of network formation. Our analysis presumes the researcher observes only a *single* network, consisting of N agents, for a small number of time periods, T . The model features two high-dimensional nuisance parameters: (i) the $\binom{N}{2}$ dyad-specific unobserved ‘incidental’ parameters $\mathbf{A} = [A_{ij}]_{1 \leq i \neq j \leq N}$, which capture heterogeneity in the payoff to link-formation across pairs of agents, and (ii) the probability mass function for the initial network (which assigns probabilities to the $2^{\binom{N}{2}}$ possible initializations of the observed sequence of networks).

Learning from a single network sequence is challenging. The presence of cross-dyad dependence in link formation, both for structural reasons and via (unmodeled) dependence across the elements of $\mathbf{A}$, complicates identification, estimation and inference. The likelihood for the observed network sequence does not naturally factor into a product of independent multiplicands; how information grows, for example, with the number of agents, N , is not readily apparent (cf., Goldsmith-Pinkham and Imbens, 2013). Although our results are specific to a particular data generating process, we expect that several of the methods introduced below could be adapted to analyze other models of dynamic network formation (in fixed T settings).

We proceed as follows. Section 2 derives a conditional likelihood for the observed network sequence that is invariant to both the dyad-specific heterogeneity and initial network condition. This likelihood is complicated; posing both computational and conceptual challenges. We therefore consider a more restrictive conditioning event. The associated conditional likelihood in this case *is* computationally and analytically tractable. With some work, it can be factored into a product of conditionally independent multiplicands. This facilitates likelihood evaluation, as well as formal statistical analysis.

Section 3, building on the results of Section 2, introduces two approaches to conducting infer-

ence on the common structural parameters – namely the coefficients indexing, respectively, the strength of structural state-dependence, α , and agents’ taste for transitivity, β . The first approach is *exact*, as in Fisher (1922). We propose a simple Metropolis-Hasting algorithm for enumeration of the elements of the conditioning set (and hence null distribution simulation). The second approach involves an asymptotic approximation as the number of agents in the *single* observed network sequence grows large. Information accumulation in this setting is delicate. An upshot of our formal analysis is that it provides insight into what types of real-world settings our methods are *a priori* mostly likely to be of value.

We further explore whether and how information accumulates as the network grows via two examples. In the first example the ($t = 0$) initial network is an Erdős-Renyi random graph. The network evolves in subsequent periods ($t = 1, 2, 3$) according to our dynamic model. We assume that A_{ij} lies in a bounded interval for all dyads. In this setting, we show that the probability of our conditioning event shrinks to zero as N grows large. Consistent estimation is not possible, although informative finite sample inference may be.

In a second example, we use a random geometric graph structure (see, for example, Penrose (2003)) to impose a strong level of homophily on unobservables. In this example, any pair of agents that are “far enough” apart never link. We show that information *does* accumulate with N in this setting, supporting consistent estimation.

While our identification argument imposes no restrictions on the distribution of unobserved heterogeneity, a researcher’s ability to consistently estimate model parameters, as these examples illustrate, does vary with features of this distribution. This is an old, albeit under-appreciated, observation in the context of single-agent panel data analysis (see, for example, Andersen (1970)). In our more complicated multi-agent network setting, the heterogeneity distribution appears to more overtly limit statistical learning.

Section 4 explores the finite sample properties of our methods via a series of Monte Carlo experiments. These experiments closely track the predictions of our asymptotic theory. In Section 5 we revisit the co-authorship network compiled by Ductor et al. (2014b). As a “proof-of-concept” we use these data, and our methods, to explore the initiation and dissolution of co-authorship among researchers.

In follows we interchangeably use unit, node, vertex, agent and individual to refer to the $i = 1, \dots, N$ vertices of the sampled network or graph. We denote random variables by capital Roman letters, specific realizations by lower case Roman letters, and their support by blackboard bold Roman letters. That is Y , y and $\mathbb{Y}$ respectively denote a generic random draw of, a specific value of, and the support of, Y . In general, bold-facing, denotes a matrix-valued random variable, $\mathbf{D}$, or its sample realization, $\mathbf{d}$. We also use the shorthand $\Pr(y|x) = \Pr(Y = y|X = x)$ throughout. As is standard in the literature on U-Statistics, we use the

shorthand $\sum_{i < j}$ for $\sum_{i=1}^{N-1} \sum_{j=i+1}^N$.

1 A dynamic network formation model

Let $\mathcal{V} = \{1, \dots, N\}$ be the set of vertices (agents) on which a graph of undirected edges $\mathcal{E}_t \subseteq \mathcal{V} \times \mathcal{V} \setminus \Delta(\mathcal{V})$ forms.² We assume the graph is observed in each of periods $t = 0, 1, \dots, T$. The adjacency matrix $\mathbf{D}_t \stackrel{\text{def}}{=} [D_{ijt}]_{1 \leq i, j \leq N}$ records the presence or absence of all possible edges in the period t graph $G_t = \mathcal{G}(\mathcal{V}, \mathcal{E}_t)$. This matrix is symmetric, with zeroes on its diagonal (since self-loops are not allowed). We let $\mathbb{D}_N$ denote the set of all $2^{\binom{N}{2}}$ possible $N \times N$ adjacency matrices.

Our analysis leaves the initial condition of the network, $\mathbf{D}_0$, unmodeled. For periods $t = 1, \dots, T$ we assume that agents add, subtract, or maintain links in order to maximize a per-period payoff function. Let $\nu_i : \mathbb{D}_N \rightarrow \mathbb{R}$ denote agent i 's payoff function. This function maps adjacency matrices – equivalently networks – into the hedonimeter-measurable utils of Francis Ysidro Edgeworth, as described in his seminal book *Mathematical Psychics*,

$$\nu_i \left(\mathbf{d}_t | \mathbf{D}_{t-1}, \tilde{\mathbf{U}}_t; \tilde{\mathbf{A}}; \theta_0 \right) = \sum_j d_{ijt} \left[\frac{\alpha_0}{2} D_{ijt-1} + \frac{\beta_0}{2} \left(\sum_k D_{ikt-1} D_{jkt-1} \right) + \tilde{A}_{ij} - \tilde{U}_{ijt} \right], \quad (2)$$

with $\tilde{\mathbf{U}}_t \stackrel{\text{def}}{=} [\tilde{U}_{ijt}]_{1 \leq i \neq j \leq N}$ a matrix of *time-varying* random utility shifters, $\tilde{\mathbf{A}} \stackrel{\text{def}}{=} [\tilde{A}_{ij}]_{1 \leq i \neq j \leq N}$ a matrix of *time-invariant* dyad-specific heterogeneity, and $\theta = (\alpha, \beta)' \in \Theta$ the target parameter of interest. While agents observe $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{U}}_t$, the econometrician does not. The goal is to learn the value of θ from the observed sequence of networks $\mathbf{D}_0, \mathbf{D}_1, \dots, \mathbf{D}_T$.

Specification (2) implies that i attaches greater utility to a link with j in period t if (i) they were connected in the prior period and (ii) they shared friends in common in the prior period (the term $\sum_k D_{ikt-1} D_{jkt-1}$ equals a count of the number of agents k which were connected to both i and j in period $t-1$).³

We assume that the random utility *pairs* $\left\{ \tilde{U}_{ijt}, \tilde{U}_{jit} \right\}_{1 \leq i < j \leq N, t=1, \dots, T}$ are independently and identically distributed across dyads as well as over time. Below we make additional parametric assumptions on this distribution. In contrast we treat $\tilde{\mathbf{A}}$ as a random matrix whose distribution is left nonparametric.^{4,5} Consequentially, our setup accommodates complex forms of

²Here $\Delta(\mathcal{V})$ represents the diagonal set of disallowed self-loops

³For simplicity we assume that $\alpha > 0$ and $\beta > 0$ during the exposition, although this is not a restriction of the model.

⁴Much of our analysis follows when $\tilde{\mathbf{A}}$ is instead viewed as a nonstochastic (and very high-dimensional) parameter.

⁵Although we place no formal restrictions on the distribution of $\mathbf{A}$, its realized value will influence the

dependence across the elements of $\tilde{\mathbf{A}}$. It may be, for example, that agent i is an extrovert and generically enjoys connecting with others, such that $\tilde{A}_{ij}$ and $\tilde{A}_{ik}$ positively co-vary. Alternatively, if i, j and k all share a latent attribute upon which agents homophilously sort, then $\tilde{A}_{ij}$, $\tilde{A}_{ik}$ and $\tilde{A}_{jk}$ will co-vary. Dyadic “clustering” of this type arises frequently in network settings (Fafchamps and Gubert, 2007; Graham, 2020).

The marginal utility received by agent i when edge ij is added to, or subtracted from, $\mathbf{D}_t$ equals

$$MU_{ij}(\mathbf{D}_t) = \begin{cases} \nu_i(\mathbf{D}_t) - \nu_i(\mathbf{D}_t - ij) & \text{if } D_{ijt} = 1 \\ \nu_i(\mathbf{D}_t + ij) - \nu_i(\mathbf{D}_t) & \text{if } D_{ijt} = 0 \end{cases}$$

where $\mathbf{D}_t - ij$ is the adjacency matrix associated with the network obtained after deleting edge ij and $\mathbf{D}_t + ij$ the one obtained via link addition (and we suppress the dependence of ν_i on terms other than $\mathbf{D}_t$)

We assume that utility is transferable across agents, such that i and j may freely share any surplus generated by an edge between them.⁶ This assumption allows us to invoke the pairwise stability with transfers equilibrium concept of Bloch and Jackson (2006).

Definition 1. (PAIRWISE STABILITY WITH TRANSFERS) The network $G(\mathcal{V}, \mathcal{E}_t)$ is pairwise stable with transfers if

- (i) $\forall ij \in \mathcal{E}_t(G), MU_{ij}(\mathbf{D}_t) + MU_{ji}(\mathbf{D}_t) \geq 0$
- (ii) $\forall ij \notin \mathcal{E}_t(G), MU_{ij}(\mathbf{D}_t) + MU_{ji}(\mathbf{D}_t) < 0$

Pairwise stable networks are those where any dyad ij , on net, weakly benefits from edge D_{ijt} when present and *would not* benefit from adding it to the network when not present (otherwise the dyad could form the edge and split the surplus).

When the utility function is of the form given in (2) the marginal utility agent i gets from a link with j equals

$$MU_{ij}(\mathbf{d}_t) = \frac{\alpha_0}{2} D_{ijt-1} + \frac{\beta_0}{2} \left(\sum_k D_{ikt-1} D_{jkt-1} \right) + \tilde{A}_{ij} - \tilde{U}_{ijt}.$$

Definition 1 then implies that – conditional on $\tilde{\mathbf{A}}$, the realized values $\left\{ \tilde{U}_{ijt}, \tilde{U}_{jit} \right\}_{1 \leq i < j \leq N}$, and

informativeness of our exact inference methods, and its underlying distribution the rate at which information accumulates in our asymptotic analysis. Similar issues arise in single agent binary choice panel data models. See, for example, Andersen (1970) and Chamberlain (1980).

⁶Goldsmith-Pinkham and Imbens (2013) study a dynamic network formation model with non-transferrable utility.

the magnitude of θ_0 – the networks must each satisfy, for $i = 1, \dots, N-1$ and $j = i+1, \dots, N$,

$$D_{ijt} = \mathbf{1} \left(\alpha_0 D_{ijt-1} + \beta_0 \left(\sum_k D_{ikt-1} D_{jkt-1} \right) + A_{ij} \geq U_{ijt} \right), \quad t = 1, \dots, T, \quad (3)$$

where $A_{ij} \stackrel{\text{def}}{=} \tilde{A}_{ij} + \tilde{A}_{ji}$ and $U_{ijt} \stackrel{\text{def}}{=} \tilde{U}_{ijt} + \tilde{U}_{jit}$.

Observe that since $(\tilde{U}_{ijt}, \tilde{U}_{jit})$ is distributed independently and identically across dyads and over time, so is U_{ijt} . Let $F_U(u)$ denote the distribution function of U_{ijt} , we have

$$F(U_{121}, \dots, U_{12T}, \dots, U_{N-1N1}, \dots, U_{N-1NT}) = \prod_{i < j} \prod_{t=1}^T F_U(U_{ijt}). \quad (4)$$

Throughout we assume that $F_U(u)$ is strictly increasing on $\mathbb{R}^1$.

Finally, since (3) only applies to periods $t = 1, \dots, T$ we leave the joint distribution of $(\mathbf{D}_0, \mathbf{A})$, the initial condition of the heterogeneity, unrestricted:

$$(\mathbf{D}_0, \mathbf{A}) \sim \Pi_0. \quad (5)$$

The mass-density function evaluated at $\mathbf{D}_0 = \mathbf{d}_0$, $\mathbf{A} = \mathbf{a}$ is denoted by $\pi_0(\mathbf{d}_0, \mathbf{a})$.

Some observations

Before proceeding, it is helpful to make a few observations about the network formation process described by equations (3), (4) and (5). First, the nuisance parameter space is very “large”; $\mathbf{A}$ consists of $\binom{N}{2}$ real-valued unknown heterogeneity terms, while Π_0 assigns probability mass to all possible $2^{\binom{N}{2}}$ initial network configurations. In contrast, the target parameter, θ , consists of just two elements.

Second, the network evolves according to a heterogenous Markov process.⁷

$$\Pr(\mathbf{D}_t = \mathbf{d}_t | \mathbf{D}_{t-1} = \mathbf{d}_{t-1}, \dots, \mathbf{D}_0 = \mathbf{d}_0, \mathbf{A}) = \Pr(\mathbf{D}_t = \mathbf{d}_t | \mathbf{D}_{t-1} = \mathbf{d}_{t-1}, \mathbf{A})$$

The process is Markov conditional on $\mathbf{A}$, but will generally not be unconditionally. Third, and relatedly, the network follows a type of nonlinear vector-autoregressive process; albeit one with substantial structure on how $\mathbf{D}_{t-1}$ influences $\mathbf{D}_t$.

Fourth, the model parsimoniously includes three forces long-hypothesized by researchers as important drivers of link formation (cf., Snijders, 2011). First, there is *state dependence*:

⁷While this process is heterogeneous due to $\mathbf{A}$, it is time homogenous.

agents i and j are more likely to form a link in period t if they were connected in period $t - 1$ (parameterized by α). Second, agents link more frequently when doing so generates “triadic closure”: links which “close” two-stars left open in the prior period are more likely (parameterized by β). Third, links may form because of unobserved good ‘fundamentals’ (i.e., A_{ij} is high).

One source of ‘good fundamentals’ is that the pair ij might be similar in some salient unobserved dimension. The tendency for individuals to assortatively match on various *observed* characteristics is well documented (McPherson et al., 2001). Here A_{ij} might reflect utility from assortative matching on *unobserved* attributes. To make this idea concrete let ξ_i be a vector of latent agent-specific characteristics and $g(\xi_i, \xi_j)$ a measure of the distance between i and j (i.e., $g(\cdot, \cdot)$ is a distance function). If $A_{ij} = -g(\xi_i, \xi_j)$, then rule (3) implies that a link between i and j is more likely if they are similar in terms of ξ . Note A_{ij} could reflect more than just homophily. For example, setting

$$A_{ij} = \lambda_i + \lambda_j - g(\xi_i, \xi_j), \quad (6)$$

allows for the possibility that certain individuals may uniformly generate higher friendship surplus. Put differently λ_i , and hence A_{ij} , might be high because individual i is a ‘good friend’. Effects of this type give rise to *degree heterogeneity* or variation in the number of links maintained by different individuals (e.g., Krivitsky et al., 2009; Graham, 2017). Degree heterogeneity is also a feature of many real-world networks.

Both a *structural* taste for transitivity in relationships, here parameterized by β_0 , and homophily on unobserved attributes, captured by $\mathbf{A}$, can generate high levels of clustering in networks. Heuristically, this is why identification and estimation are challenging in our setting.

Although we utilize an equilibrium notion drawn from game theory (Definition 1 above), agents in our model – while maximizing utility – are not playing a game (at least a non-trivial one). In each period, agents condition their actions on the structure of the network in the *prior* period and add, subtract, or maintain links myopically. Extending the work reported here to accommodate forward-looking agents, and the game-theoretic difficulties that doing so would raise, merits exploration.

As an alternative to specifying a dynamic game, one might ask what can be learned from a single network observation. Pelican and Graham (2026) and Graham and Pelican (2020) study static, complete information, games of network formation in the presence of agent-level heterogeneity. While they report positive results, doing so requires much stronger assumptions on the structure of unobserved heterogeneity, $\mathbf{A}$, than we maintain here. Without

restrictions on $\mathbf{A}$, *any* network configuration can be generated by an appropriate draw of $\mathbf{A}$. Our results circumvent this problem by exploiting the availability of multiple network observations over time and the assumed *time-invariance* of $\mathbf{A}$.

Likelihood

Equations (3), (4) and (5) specify a dynamic model of network formation. To ease the notation, let $R_{ijt} \stackrel{\text{def}}{=} \sum_k D_{ikt} D_{jkt}$. The joint probability mass-density of a specific sequence of network configurations $\mathbf{D}^T \stackrel{\text{def}}{=} (\mathbf{D}_0, \mathbf{D}_1, \dots, \mathbf{D}_T)$ and realization of $\mathbf{A}$ is

$$\begin{aligned} p(\mathbf{d}^T, \mathbf{a}; \theta, \pi) &= \pi(\mathbf{d}_0, \mathbf{a}) \\ &\times \prod_{i < j} \prod_{t=1}^T F(\alpha d_{ijt-1} + \beta r_{ijt-1} + a_{ij})^{d_{ijt}} \\ &\times [1 - F(\alpha d_{ijt-1} + \beta r_{ijt-1} + a_{ij})]^{1-d_{ijt}}. \end{aligned} \quad (7)$$

One approach to estimating θ would involve (i) assuming F_U is known (e.g., $F_U(u) = \Phi(u)$ with $\Phi(u)$ the CDF of a standard normal random variable) and (ii) additionally parameterizing $\pi(\mathbf{d}_0, \mathbf{a})$. This is the approach taken by Goldsmith-Pinkham and Imbens (2013) in their own *Journal of Business and Economic Statistics* lecture over 10 years ago (a lecture that directly inspired the beginning of the research reported in this paper). Let $\pi_{\mathbf{D}_0|\mathbf{A}}(\mathbf{d}_0|\mathbf{a}; \eta_1)$ denote a parametric family, indexed by η_1 , for $\Pr(\mathbf{D}_0 = \mathbf{d}_0 | \mathbf{A} = \mathbf{a})$, the distribution of the initial network given $\mathbf{A} = \mathbf{a}$. Let $\pi_{\mathbf{A}}(\mathbf{a}; \eta_2)$ denote a parametric family, indexed by η_2 , for the marginal density function for $\mathbf{A}$. The integrated likelihood for the observed data is then

$$\begin{aligned} p^l(\mathbf{d}^T; \theta, \eta) &= \int \dots \int \prod_{t=1}^T \prod_{i < j} \left\{ F_U(\alpha d_{ijt-1} + \beta r_{ijt-1} + a_{ij})^{d_{ijt}} \right. \\ &\quad \times [1 - F_U(\alpha d_{ijt-1} + \beta r_{ijt-1} + a_{ij})]^{1-d_{ijt}} \Big\} \\ &\quad \times \pi_{\mathbf{D}_0|\mathbf{A}}(\mathbf{d}_0|\mathbf{a}; \eta_1) \pi_{\mathbf{A}}(\mathbf{a}; \eta_2) da_{21}, \dots, da_{NN-1} \end{aligned} \quad (8)$$

with $\eta = (\eta'_1, \eta'_2)'$. Although (8) represents a parametric family for the observed data, using it as a basis for estimation and inference raises at least three conceptual and computational issues.

First, as is familiar from prior work on dynamic discrete choice analysis, rule (3) provides no guidance on how to specify $\pi_{\mathbf{D}_0|\mathbf{A}}(\mathbf{d}_0|\mathbf{a}; \eta_1)$ (Heckman, 1981c; Honoré and Tamer, 2006). In analogy to its panel counterpart, we call this the *initial network problem*. Rule (3) does suggest that the probability of the event $\mathbf{D}_0 = \mathbf{d}_0$ should vary with the realized value of $\mathbf{A}$, but

little else. One approach, again inspired by single agent models, would involve assuming that $\pi_{\mathbf{D}_0|\mathbf{A}}(\mathbf{d}_0|\mathbf{a}, \eta_1)$ coincides with the steady state distribution implied by (3) (e.g., Heckman, 1981a; Card and Sullivan, 1988). Even if this is empirically plausible, operationalizing it would be non-trivial.

A second problem is that, even if $\pi_{\mathbf{D}_0|\mathbf{A}}(\mathbf{d}_0|\mathbf{a}, \eta_1)$ were correctly specified, evaluating (8) requires computing a very high-dimensional integral. Since it seems reasonable, at the very minimum, to choose a specification for $\pi_{\mathbf{A}}(\mathbf{a}; \eta_2)$ which allows A_{ij} and A_{ik} to covary, there is no obvious way to factor (8) to reduce the dimensionality of the required integration. Goldsmith-Pinkham and Imbens (2013) assume that $A_{ij} = \alpha_\xi |\xi_i - \xi_j|$ with ξ_i binary, independent of ξ_j , and $\Pr(\xi_i = 1) = p$. Directly evaluating the integrated likelihood in this case involves a weighted sum of the likelihood given $\mathbf{A}$ over its 2^N possible realizations.⁸

A third problem, related to the second, and emphasized by Goldsmith-Pinkham and Imbens (2013), is that even if the maximum likelihood estimate could be computed, it would not be clear how to conduct large sample inference using $(\hat{\alpha}_{ML}, \hat{\beta}_{ML})$; at least with data drawn from a single network. This motivates their recourse to Bayesian methods, which are attractive for computational reasons and also for providing a principled approach to inference. The main approach to inference developed below, in contrast, is frequentist (albeit one that involves a delicate asymptotic argument).

2 Conditional likelihood analysis

While integrated likelihood analysis merits further exploration, our focus will instead lie with methods which remain valid irrespective of the precise form of π and/or realized configuration of $\mathbf{A}$: so called *fixed effects* methods.

Hereon we maintain a logistic assumption on the $\{U_{ijt}\}_{i < j, t=1, \dots, T}$.⁹ For simplicity we also focus on the $T = 3$ case.¹⁰ Under the logistic assumption, reminiscent of Chamberlain

⁸van Duijn et al. (2004) introduce MCMC methods that also might be adapted in order to evaluate (8) in special cases.

⁹That is, the CDF of U_{ijt} equals $\Pr(U_{ijt} \leq u) = F_U(u) = \frac{\exp(u)}{1 + \exp(u)}$.

¹⁰By aggregating across (possibly overlapping) sub-panels of length four, we can accommodate longer network sequences. However, fully exploiting the identifying content of longer panels would require substantively extending the results which follow.

(1985), we can write the likelihood of the event $\mathbf{D}^3 = \mathbf{d}^3$ as

$$\begin{aligned} \Pr(\mathbf{D}^3 = \mathbf{d}^3; \theta, \mathbf{A}, \pi) &= \pi(\mathbf{d}_0; \mathbf{A}) \\ &\times \prod_{i < j} \prod_{t=1}^3 \left[\frac{\exp(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})}{1 + \exp(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})} \right]^{d_{ijt}} \\ &\times \left[\frac{1}{1 + \exp(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})} \right]^{1-d_{ijt}} \\ &= \frac{\pi(\mathbf{d}_0; \mathbf{A}) \times \prod_{i < j} \exp(\alpha \sum_{t=1}^3 d_{ijt-1} d_{ijt} + \beta \sum_{t=1}^3 r_{ijt-1} d_{ijt} + A_{ij} \sum_{t=1}^3 d_{ijt})}{\prod_{i < j} \prod_{\delta=0,1} \prod_{\rho=0}^{N-2} [1 + \exp(\alpha \delta + \beta \rho + A_{ij})]^{m_{ij}(\delta, \rho; \mathbf{d}^3)}}, \quad (9) \end{aligned}$$

with

$$m_{ij}(\delta, \rho; \mathbf{d}^3) = |\{t = 1, 2, 3 : d_{ij,t-1} = \delta, r_{ij,t-1} = \rho\}|$$

equal to the within- ij -dyad count of the number of link decisions d_{ijt} preceded by different combinations of feasible values for $d_{ijt-1} \in \{0, 1\}$ and $r_{ijt-1} \in \{0, 1, \dots, N-2\}$. For example, $m_{ij}(0, 3)$ counts the number of times dyad ij considers forming a link when, in the period prior, they were (i) unlinked and (ii) shared 3 friends in common.

Recall that $\mathbb{D}_N$ denotes the set of $N \times N$ undirected binary adjacency matrices. Let $\mathbb{D}_N^3 \stackrel{\text{def}}{=} (\mathbb{D}_N, \mathbb{D}_N, \mathbb{D}_N, \mathbb{D}_N)$ denote the set of all possible length $T = 3$ sequences of such networks, and consider the restricted set of network sequences:

$$\begin{aligned} \mathbb{B}^*(\mathbf{d}^3) &= \{ \mathbf{b}^3 = (\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) : \mathbf{b}_0 = \mathbf{d}_0, \\ &\quad \forall i = 1, \dots, N-1, j = i+1, \dots, N, \\ &\quad \& \forall \delta = 0, 1 \& \rho = 0, 1, \dots, N-2, \\ &\quad \sum_{t=1}^3 b_{ijt} = \sum_{t=1}^3 d_{ijt}, \\ &\quad \& m_{ij}(\delta, \rho; \mathbf{b}^3) = m_{ij}(\delta, \rho; \mathbf{d}^3), \\ &\quad \& b_{ijt} = b_{jit} \in \{0, 1\} \text{ for } t = 0, 1, 2, 3 \}. \end{aligned} \quad (10)$$

The denominator of $\Pr(\mathbf{D}^3 = \mathbf{b}^3; \theta, \mathbf{A}, \pi)$ coincides with that appearing in (9) for all $\mathbf{b}^3 \in \mathbb{B}^*(\mathbf{d}^3)$. The initial condition, π , as well as the contributions of the A_{ij} appearing in the numerator of (9), additionally coincide. These observations deliver the conditional probability:

$$\Pr(\mathbf{d}^3 | \mathbf{d}^3 \in \mathbb{B}^*; \theta, \mathbf{A}, \pi) = \frac{\prod_{i < j} \exp(\alpha \sum_{t=1}^3 d_{ijt-1} d_{ijt} + \beta \sum_{t=1}^3 r_{ijt-1} d_{ijt})}{\sum_{\mathbf{b} \in \mathbb{B}^*(\mathbf{d}^3)} \prod_{i < j} \exp(\alpha \sum_{t=1}^3 b_{ijt-1} b_{ijt} + \beta \sum_{t=1}^3 s_{ijt-1} b_{ijt})}, \quad (11)$$

where we let $s_{ijt} \stackrel{\text{def}}{=} \sum_k b_{ikt} b_{jkt}$. The right-hand-side of (11) does not vary with $\mathbf{A}$ or $\pi(\mathbf{d}_0; \mathbf{A})$; we have found an implication of the model, indexed by θ , that is free of the nuisance parameters.

The set $\mathbb{B}^*(\mathbf{d}^3)$ appears rather complicated. Indeed, a superficial glance does not indicate whether its cardinality is large, and hence using (11) for inference on α and β likely informative, or whether “conditioning away” the high-dimensional nuisance parameters $\mathbf{A}$ and π leaves effectively no identifying variation to work with. Furthermore, (13) suggest no obvious independence structure indicating how, or even if, information might accumulate as the number of agents, N , in the network grows large.

Our first result, Proposition 1 immediately below, provides an alternative characterization of $\mathbb{B}^*(\mathbf{d}^3)$; a characterization that facilitates the derivation of several results presented below. Defining (see below for commentary)

$$\begin{aligned} \mathbb{B}(\mathbf{d}^3) = \{ \mathbf{b}^3 = (\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) : & \mathbf{b}_0 = \mathbf{d}_0, \mathbf{b}_3 = \mathbf{d}_3, \\ & \mathbf{b}_1 + \mathbf{b}_2 = \mathbf{d}_1 + \mathbf{d}_2, \forall i = 1, \dots, N-1, j = i+1, \\ & (s_{ij1}, s_{ij2}) = (r_{ij1}, r_{ij2}) \text{ if } (b_{ij1}, b_{ij2}) = (d_{ij1}, d_{ij2}) \\ & \& (s_{ij1}, s_{ij2}) = (r_{ij2}, r_{ij1}) \text{ if } (b_{ij1}, b_{ij2}) = (d_{ij2}, d_{ij1}), \\ & \& b_{ijt} = b_{jit} \in \{0, 1\} \text{ for } t = 0, 1, 2, 3 \}, \end{aligned} \quad (12)$$

allows us to state:

Proposition 1. (SUFFICIENT CONDITIONING SET)

- (i) $\mathbf{d}_0$, $\sum_{t=1}^3 d_{ijt}$, and $m_{ij}(\delta, \rho; \mathbf{d}^3)$ for $\forall i = 1, \dots, N-1, j = i+1, \dots, N$ are sufficient statistics for $\mathbf{A}$ and π ;
- (ii) $\mathbb{B}^*(\mathbf{d}^3) = \mathbb{B}(\mathbf{d}^3)$.

Proof. See Appendix B.

Proposition 1 gives a set of sufficient statistics for $\mathbf{A}$ and π . It also provides a convenient alternative characterization of the set of all network sequences where these sufficient statistics coincide with those found in the observed network sequence, $\mathbf{d}^3$.

Using Proposition 1 – see also Dano et al. (2025) – we can rewrite the conditional likelihood

(11) as:

$$\begin{aligned} \Pr(\mathbf{d}^3 | \mathbf{d}^3 \in \mathbb{B}(\mathbf{d}^3); \theta, \mathbf{A}, \pi) &= \Pr(\mathbf{d}^3 | \mathbf{d}^3 \in \mathbb{B}(\mathbf{d}^3); \theta) \\ &= \frac{\exp\left(\alpha \sum_{i < j} \sum_{t=1}^3 d_{ijt-1} d_{ijt} + \beta \sum_{i < j} \sum_{t=1}^3 r_{ijt-1} d_{ijt}\right)}{\sum_{\mathbf{b} \in \mathbb{B}(\mathbf{d}^3)} \exp\left(\alpha \sum_{i < j} \sum_{t=1}^3 b_{ijt-1} b_{ijt} + \beta \sum_{i < j} \sum_{t=1}^3 s_{ijt-1} b_{ijt}\right)}. \end{aligned} \quad (13)$$

The likelihood (13), by conditioning on sufficient statistics for $\mathbf{A}$ and π , generates a probability law for $\mathbf{D}^3 | \mathbf{D}^3 \in \mathbb{B}(\mathbf{d}^3)$ indexed only by the finite-dimensional parameter θ . For a given population value of $\theta = \theta_0$, equation (13) assigns different *ex ante* probabilities to network sequences in $\mathbb{B}(\mathbf{d}^3)$. The (conditional) maximum likelihood principle suggests choosing $\hat{\theta}$ to make this *ex ante* probability maximal for the realized network sequence, $\mathbf{d}^3$.

Although free of nuisance parameters, using (13) as a basis for estimation of, and inference on, θ is not straightforward. Evaluating the denominator of (13) requires enumerating the elements of $\mathbb{B}(\mathbf{d}^3)$, a set with non-trivial geometry and potentially intractably large cardinality.

Although exact evaluation of (13) appears challenging, it may be possible to construct a simulation-based estimate of it. Doing so requires a method of sampling from $\mathbb{B}(\mathbf{d}^3)$ uniformly at random. The observed network sequence $\mathbf{D}^3 = \mathbf{d}^3$ is an element of $\mathbb{B}(\mathbf{d}^3)$ by construction. To find a new element in $\mathbb{B}(\mathbf{d}^3)$ we can randomly rewire $\mathbf{d}^3$.

We call dyads where $(d_{ij1}, d_{ij2}) = (0, 1)$ or $(d_{ij1}, d_{ij2}) = (1, 0)$ *movers*, since they change their link status between periods 1 and 2. We say mover ij 's edges are “swapped” in $\mathbf{b}^3$ relative to $\mathbf{d}^3$ if $(b_{ij1}, b_{ij2}) = (d_{ij2}, d_{ij1})$. Since the initial and terminal conditions of $\mathbf{b}^3$ are constant across all $\mathbf{b}^3 \in \mathbb{B}(\mathbf{d}^3)$ and, furthermore, $b_{ij1} + b_{ij2} = d_{ij1} + d_{ij2}$ for all dyads ij , a move from $\mathbf{d}^3$ to *any* $\mathbf{b}^3 \in \mathbb{B}(\mathbf{d}^3)$ can be engineered by swapping edges for a subset of *mover* dyads.

Implementing swaps of this type, however, may induce violations of the constraint that $(s_{ij1}, s_{ij2}) = (r_{ij1}, r_{ij2})$ if $(b_{ij1}, b_{ij2}) = (d_{ij1}, d_{ij2})$ and $(s_{ij1}, s_{ij2}) = (r_{ij2}, r_{ij1})$ if $(b_{ij1}, b_{ij2}) = (d_{ij2}, d_{ij1})$. To understand why, observe that d_{ijt} is a component of r_{ikt} for $k \neq i, j$ and similarly for r_{jkt} (for $k \neq i, j$). Hence, if b_{ij1} and b_{ij2} are a swap of the observed d_{ij1} and d_{ij2} , then up to $2 \times 2(N - 2)$ of the “friends-in-common” regressor values in $\mathbf{b}^3$ may differ from their corresponding values in $\mathbf{d}^3$. The set $\mathbb{B}(\mathbf{d}^3)$ allows for such differences, but only of a very particular form. Specifically, it requires that if b_{ij1} and b_{ij2} are a swap of the observed d_{ij1} and d_{ij2} , then it must be the case that s_{ij1} and s_{ij2} are also a swap of the observed r_{ij1} and r_{ij2} .

Consequently, successfully moving from one element in $\mathbb{B}(\mathbf{d}^3)$ to another may require the simultaneous swapping of *many* movers' edges to ensure that any constraint violations induced

by one group of edge swaps is appropriately offset by another.

A Markov Chain Monte Carlo (MCMC) algorithm for constrained graph simulation introduced by Pelican and Graham (2026) provides a template for how one might approach this problem. We do not pursue this approach. Instead, we will characterize a subset of $\mathbb{B}(\mathbf{d}^3)$, that we *can* explicitly enumerate and develop corresponding conditional estimators. Our approach, since it conditions on only a subset of those network sequences with common sufficient statistics for $\mathbf{A}$ and $\pi(\mathbf{d}_0; \mathbf{A})$, involves some loss of information. Its virtues are analytical and computational tractability.

Star systems

In this subsection we define a *subset* of $\mathbb{B}(\mathbf{d}^3)$ whose elements can be enumerated by simultaneously swapping all period 1 and period 2 edges among a set of dyads which we will call *identifying star systems*. An identifying star system is a set of dyads such that, if we swap all its period 1 edges with their period 2 counterparts, then we move to a new network sequence that is *guaranteed* to remain in $\mathbb{B}(\mathbf{d}^3)$. Identifying star systems provide a simple way to move through a (subset) of the elements of $\mathbb{B}(\mathbf{d}^3)$. Star systems vary in the number of dyads constituting them. We show how to utilize all those systems composed of $p = 2 \dots K$ agents. Let $\mathbb{B}_K(\mathbf{d}^3) \subset \mathbb{B}(\mathbf{d}^3)$ denote the set of all network sequences reachable by swapping the edges of identifying star systems of size K or smaller (defined formally below). If the observed network sequence, $\mathbf{d}^3$, contains $\mathbf{m}_{2:K,N}$ identifying star systems of size K or smaller, then $|\mathbb{B}_K(\mathbf{d}^3)| = 2^{\mathbf{m}_{2:K,N}}$, corresponding to all the possible ways in which the $\mathbf{m}_{2:K,N}$ identifying K -star systems may have their edges swapped in $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ relative to their configuration in $\mathbf{d}^3$ or not. We say a star system's edges are "swapped" in $\mathbf{b}^3$ if they are permuted relative to their configuration in the observed network sequence $\mathbf{d}^3$.

In what follows we explore estimators which condition on the event $\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)$. Considering the conditional probability of observing the network sequence in hand, given that it belongs to a set of network sequences that can be enumerated by swapping the edges of identifying star systems, provides an analytically tractable conditioning event and an associated criterion function amenable to asymptotic analysis.

To describe our estimator and characterize its sampling properties we need to first carefully describe $\mathbb{B}_K(\mathbf{d}^3)$.

Definition 2. (P-STAR SYSTEM) The dyad set $\mathcal{S}_p = \{i_1 i_2, i_1 i_3, \dots, i_1 i_p\}$ corresponds to a *p-star system* with central vertex i_1 . Let $v(\mathcal{S}_p) = \{i_1, \dots, i_p\}$ denote the agents belonging to $\mathcal{S}_p$.

Definition 3. (NEIGHBORHOOD) The period t *neighborhood* of $\mathcal{S}_p$ includes all agents not in $v(\mathcal{S}_p)$ but connected to an agent in this set during period t :

$$n_t(\mathcal{S}_p) = \{j : d_{jit} = 1 \text{ for a least one } i \in v(\mathcal{S}_p)\} \setminus \{i_1, \dots, i_p\}. \quad (14)$$

Definition 4. (STABLE NEIGHBORHOOD) The neighborhood of $\mathcal{S}_p = \{i_1 i_2, i_1 i_3, \dots, i_1 i_p\}$ is *stable* if

- (i) $\forall i \in v(\mathcal{S}_p)$ and $j \in \{1, \dots, N\} \setminus v(\mathcal{S}_p)$: $d_{ij1} = d_{ij2}$;
- (ii) $\forall j \in n_1(\mathcal{S}_p)$ and $l \in \{1, \dots, N\}$: $d_{jl1} = d_{jl2}$;
- (iii) $\forall i_k i_l$ with $2 \leq k, l \leq p$ and $k \neq l$ $d_{i_k i_l 1} = d_{i_k i_l 2}$.

A p-star system is embedded in a stable neighborhood if all agents not in $v(\mathcal{S}_p)$, but connected to someone in $v(\mathcal{S}_p)$, maintain a fixed link structure in periods 1 and 2. This includes edges into $v(\mathcal{S}_p)$, as well as edges to other agents in the network (Parts (i) and (ii) of Definition 4). We also require that any edges among agents in $v(\mathcal{S}_p)$ that are not among the dyads in $\mathcal{S}_p$ remain fixed across periods 1 and 2 (Part (iii) of Definition 4). The joint impact of these three conditions is that, other than any edges among the dyads in $\mathcal{S}_p$, the network structure up to two degrees away from any agent in $v(\mathcal{S}_p)$ is stable across periods 1 and 2.

While we require *stability* of those links surrounding $\mathcal{S}_p$, we require *movement* of the links within it.

Definition 5. (IDENTIFYING P-STAR SYSTEM) $\mathcal{S}_p = \{i_1 i_2, i_1 i_3, \dots, i_1 i_p\}$ is *identifying* if

- (i) $\mathcal{S}_p$ is embedded in a stable neighborhood,
- (ii) $\forall i_1 i_k$ with $2 \leq k \leq p$ $d_{i_1 i_k 1} \neq d_{i_1 i_k 2}$.

Observe that if we remove one agent from the identifying p -star system $\mathcal{S}_p$, the remaining $p - 1$ agents do not constitute an identifying $(p - 1)$ -star system (condition (i) of Definition 4 is violated). Conversely, if $\mathcal{S}_p$ is *not* identifying, there may be subsets of agents in $v(\mathcal{S}_p)$ which *do* form an identifying system.

Lemma 1 and a corollary of it, as well as Lemma 2, which immediately follows, play important roles in helping us analyze the conditional likelihood $\Pr(\mathbf{D}^3 = \mathbf{d}^3 | \mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3); \theta)$.

Lemma 1. (AGENT SEPARATION) *Any node $i \in \mathcal{V}$ belongs to at most one identifying star system.*

Proof. Towards a contradiction, suppose that $i \in v(\mathcal{S}_p) \cap v(\mathcal{S}_l)$ where $\mathcal{S}_p$ and $\mathcal{S}_l$ are two distinct identifying star systems.

Case 1: If i is the central node of $\mathcal{S}_p$ and $\mathcal{S}_l$. Since $\mathcal{S}_p \neq \mathcal{S}_l$, either $\exists k \in v(\mathcal{S}_p) \setminus v(\mathcal{S}_l)$ or

$\exists k \in v(\mathcal{S}_l) \setminus v(\mathcal{S}_p)$ and part (ii) of Definition 5 implies $d_{ik1} \neq d_{ik2}$. This change in links status of dyad ik violates the assumption that $\mathcal{S}_l$ or $\mathcal{S}_p$ is embedded in a stable neighborhood.

Case 2: If i is the central node of $\mathcal{S}_p$ but not $\mathcal{S}_l$. Let j denote the center of $\mathcal{S}_l$. If $j \notin v(\mathcal{S}_p)$, then part (i) of Definition 4 implies that $d_{ij1} = d_{ij2}$. However, since j is the center node of $\mathcal{S}_l$, this contradicts part (ii) of Definition 5 which requires $d_{ij1} \neq d_{ij2}$. Hence, $j \in v(\mathcal{S}_p)$ which means that j must be a leaf node of $\mathcal{S}_p$. There are two cases. Either $\mathcal{S}_p = \mathcal{S}_l = \{ij\}$ contradicting the premise that $\mathcal{S}_p$ and $\mathcal{S}_l$ are two distinct identifying star systems. Or, without loss of generality, there exists $k \in v(\mathcal{S}_p) \setminus \{i, j\}$ and $d_{ik1} \neq d_{ik2}$ by part (ii) of Definition 5. If $k \notin v(\mathcal{S}_l)$, this contradicts part (i) of Definition 4. If $k \in v(\mathcal{S}_l)$, then by part (ii) of Definition 5, $d_{jk1} \neq d_{jk2}$. However, j and k are both leaves of $\mathcal{S}_p$, so part (iii) of Definition 4 requires $d_{jk1} = d_{jk2}$, a contradiction.

Case 3: If i is the central node of $\mathcal{S}_l$ but not $\mathcal{S}_p$, we are back to Case 2 after swapping the roles of $\mathcal{S}_l$ and $\mathcal{S}_p$.

Case 4: If i is not the central node of either $\mathcal{S}_p$ or $\mathcal{S}_l$. Let j denote the central node of $\mathcal{S}_p$. Then, we know by part (ii) of Definition 5 that $d_{ji1} \neq d_{ji2}$. Hence, if j is not the central node of $\mathcal{S}_l$, we again violate the neighborhood stability of $\mathcal{S}_l$. Finally if j is also the central node of $\mathcal{S}_l$, we are back to Case 1 with j substituting i . $\square$

Corollary 1. (DYAD SEPARATION) *Any dyad ij belongs to at most one identifying star system, hence there can be at most $\frac{N}{2}$ identifying star systems.*

Corollary 1, as we shall see, implies limits on the rate-of-convergence of our preferred estimate of θ .

Lemma 2 demonstrates that direct connections between agents in two identifying star systems are impossible.

Lemma 2. (INVARIANCE OF LINKS ACROSS IDENTIFYING STAR SYSTEMS) *Let $\mathcal{S}_p$ and $\mathcal{S}_l$ denote two distinct identifying star systems. Then, for all nodes $i \in v(\mathcal{S}_p)$ and $j \in v(\mathcal{S}_l)$, we have*

$$(A) \ D_{ij1} = D_{ij2} = 0$$

$$(B) \ R_{ij1} = R_{ij2}.$$

Proof. We begin with claim (A). Towards a contradiction, assume $D_{ij1} = 1$ or $D_{ij2} = 1$. In either case, it follows from parts (i) and (ii) of Definition 4 that $j \in n_1(\mathcal{S}_p)$ and $D_{jl1} = D_{jl2}$ for all $l = 1, \dots, N$. This violates Definition 5.(ii).

Next, for claim (B) we can write the decomposition:

$$\begin{aligned}
R_{ij2} - R_{ij1} &= \sum_{k=1}^N (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) \\
&= \sum_{k \in v(\mathcal{S}_p)} (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) + \sum_{k \in v(\mathcal{S}_l)} (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) \\
&\quad + \sum_{k \notin v(\mathcal{S}_p) \cup v(\mathcal{S}_l)} (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) \\
&= \sum_{k \in v(\mathcal{S}_p)} (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) + \sum_{k \in v(\mathcal{S}_l)} (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) + 0
\end{aligned}$$

where the second equality follows from Lemma 1 and the third equality follows from the neighborhood stability of $\mathcal{S}_p$ and $\mathcal{S}_l$. By claim (A), $\forall k \in v(\mathcal{S}_p)$, $D_{jk1} = D_{jk2} = 0$ and $\forall k \in v(\mathcal{S}_l)$, $D_{ik1} = D_{ik2} = 0$. Consequently, $R_{ij1} = R_{ij2}$ as claimed. $\square$

Lemma 2 has important implications for the type of observed network sequences that are likely to contain information about θ . Such networks will involve at least some separated clusters of links. They will not be “hairball” graphs, although they may include giant components.

Permutation Lemmas

As noted above, setting $(b_{ij1}, b_{ij2}) = (d_{ij2}, d_{ij1})$ need not result in a network sequence in $\mathbb{B}(\mathbf{d}^3)$: such swaps may induce violations of the constraint that $(s_{ij1}, s_{ij2}) = (r_{ij1}, r_{ij2})$, if $(b_{ij1}, b_{ij2}) = (d_{ij1}, d_{ij2})$, and $(s_{ij1}, s_{ij2}) = (r_{ij2}, r_{ij1})$, if $(b_{ij1}, b_{ij2}) = (d_{ij2}, d_{ij1})$.

In this section we show that swapping the period 1 and 2 edges of an identifying star system *does* result in a new network sequence which is *guaranteed* to lie in $\mathbb{B}(\mathbf{d}^3)$. We define the set $\mathbb{B}_K(\mathbf{d}^3) \subset \mathbb{B}(\mathbf{d}^3)$ to be the set of all network sequence obtainable by “swapping the edges” (or not) of identifying star systems in $\mathbf{d}^3$ of size K or less.

Lemma 3, stated and proved below, is the key tool used to establish this result. To state and prove Lemma 3 requires some additional notation and definitions. In what follows we use $\wedge$ to denote “and”, $\vee$ for “or”, and $\veebar$ for the “exclusive or”. We also use $\exists!$ to denote “there exists exactly one”.

Let $Z_{ij,p}$ denote an indicator function equal to one if dyad ij belongs to an identifying star system of size $p \geq 2$ and define the following dyad *and* vertex/agent sets.

- (i) Dyads contained in an identifying p-star system: $\mathcal{D}_p = \{ij \mid Z_{ij,p} = 1\}$.
- (ii) Agents belonging to an identifying p-star system: $\mathcal{V}_p = \{i \in \mathcal{V} \mid \exists j \in \mathcal{V} : Z_{ij,p} = 1\}$.

(iii) Dyads composed of agents in a common p-star system, but where the dyad itself is not part of the system:

$$\mathcal{D}_p^{fd} = \{ij \mid ij \notin \mathcal{D}_p \wedge i \in \mathcal{V}_p \wedge j \in \mathcal{V}_p \wedge (\exists! m \in \mathcal{V}_p : (im, jm) \in \mathcal{D}_p^2)\}.$$

Observe that for $p = 2$, $\mathcal{D}_p^{fd} = \emptyset$.

(iv) Agents belonging to some identifying star system of maximal size K : $\mathcal{V}_{2:K} = \bigcup_{p=2}^K \mathcal{V}_p$.

(v) Agents *not* belonging to an identifying star system of maximal size K : $\mathcal{V}^o = \mathcal{V} \setminus \mathcal{V}_{2:K}$.

(vi) Dyads composed of agents belonging to two distinct identifying star-systems:

$$\begin{aligned} \mathcal{D}_{k,l}^{pd} &= \{ij \mid (i \in \mathcal{V}_k \wedge j \in \mathcal{V}_l) \vee (i \in \mathcal{V}_l \wedge j \in \mathcal{V}_k)\}, \quad \text{if } l < k \\ \mathcal{D}_{k,k}^{pd} &= \{ij \mid ij \notin \mathcal{D}_k \wedge i \in \mathcal{V}_k \wedge j \in \mathcal{V}_k \wedge (\nexists m \in \mathcal{V}_k : (im, jm) \in \mathcal{D}_k^2)\}. \end{aligned}$$

(vii) Dyads with one agent in $\mathcal{V}_p$ and the other $\mathcal{V}^o$:

$$\mathcal{D}_{p,o}^{pd} = \{ij \mid (i \in \mathcal{V}_p \wedge j \in \mathcal{V}^o) \vee (i \in \mathcal{V}^o \wedge j \in \mathcal{V}_p)\}.$$

(viii) Dyads with neither agent belonging to an identifying star system of size less than K

$$\mathcal{D}_K^{od} = \{ij \mid i \in \mathcal{V}^o \wedge j \in \mathcal{V}^o\}.$$

An implication of Lemma 1 and Corollary 1, specifically the fact that an agent and a dyad can belong to at most one identifying star system, means we can partition the set of all $\binom{N}{2}$ possible dyads $\mathcal{D}$ as follows:

$$\mathcal{D} = \left(\bigcup_{p=2}^K \mathcal{D}_p \cup \mathcal{D}_p^{fd} \cup \mathcal{D}_{p,o}^{pd} \right) \cup \left(\bigcup_{p=2}^K \bigcup_{l=2}^K \mathcal{D}_{p,l}^{pd} \right) \cup \mathcal{D}_K^{od}. \quad (15)$$

Partition (15), by grouping dyads into analytically convenient subsets, simplifies likelihood analysis.

Finally, we formally define set $\mathbb{B}_K(\mathbf{d}^3)$ to consist of all network sequences obtained via permutations (i.e., “swapping the edges”) of the period $t = 1$ and $t = 2$ link decisions of identifying star systems. Permutations are applied simultaneously to all dyads in a given $\mathcal{S}_K$. All other link decisions in any $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ are held fixed at their observed values.

With the above definitions established we can state and prove the following Lemma.

Lemma 3. (PERMUTATIONS) *Define $S_{ijt} = \sum_k B_{ikt} B_{jkt}$ for $\mathbf{B}^3 \in \mathbb{B}_K(\mathbf{D}^3)$. Let $\mathcal{S}_p$ be an*

identifying p -star system in $\mathbf{D}^3$ with center i . If $(B_{ik1}, B_{ik2}) = (D_{ik1}, D_{ik2})$ for all $k \neq i$ with $k \in v(\mathcal{S}_p)$, then for each

$$(A) \quad ij \in \mathcal{S}_p, (S_{ij1}, S_{ij2}) = (R_{ij1}, R_{ij2});$$

$$(B) \quad (j, k) \in v(\mathcal{S}_p) \times v(\mathcal{S}_p) \text{ but with } jk \in \mathcal{D}_p^{fd}, (S_{jk1}, S_{jk2}) = (R_{jk1}, R_{jk2});$$

$$(C) \quad j \in v(\mathcal{S}_p) \text{ and } l \notin v(\mathcal{S}_p), (S_{jl1}, S_{jl2}) = (R_{jl1}, R_{jl2}).$$

If, instead, $(B_{ik1}, B_{ik2}) = (D_{ik2}, D_{ik1})$ for all $k \neq i$ with $k \in v(\mathcal{S}_p)$, then (A), (B) and (C) continue to hold after replacing $(R_{..1}, R_{..2})$ with $(R_{..2}, R_{..1})$.

Lemma 3 states that if, relative to their configuration in $\mathbf{d}^3$, the periods 1 and 2 edges of all dyads constituting an identifying p -star system, $\mathcal{S}_p$, are swapped, then the number of “friends-in-common” terms are also swapped. Furthermore, this swapping of “friends-in-common” extends to all dyads where at least one vertex is part of $v(\mathcal{S}_p)$. A consequence of this observation is that the likelihoods for the events $\mathbf{D}^3 = \mathbf{b}^3$ for any $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ will, after permutation, include many multiplicands-in-common. This, as we shall see, simplifies our conditional likelihood analysis.

The proof of Lemma 3 is straightforward and is provided in Appendix A.2. The idea is to repeatedly decompose the common-friends term into contributions from agents within the identifying p -star system $\mathcal{S}_p$ and from agents outside it. The implications of Definition 4 are then repeatedly applied to establish the stated results.

Conditional star system likelihood

We are now ready to define the criterion function on which our identification and estimation results rely. We begin with a formal definition of $\mathbb{B}_K(\mathbf{d}^3)$:

$$\begin{aligned} \mathbb{B}_K(\mathbf{d}^3) = & \left\{ (\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) \in \mathbb{D}_N^3 : \mathbf{b}_0 = \mathbf{d}_0, \mathbf{b}_3 = \mathbf{d}_3, \right. \\ & \text{for each identifying } p\text{-star system } \mathcal{S}_p \text{ in } \mathbf{d}^3, p = 1, \dots, K, \\ & b_{ij1} = d_{ij1} \wedge b_{ij2} = d_{ij2} \quad \forall ij \in \mathcal{S}_p \\ & \text{or } b_{ij1} = d_{ij2} \wedge b_{ij2} = d_{ij1} \quad \forall ij \in \mathcal{S}_p, \\ & \left. \text{otherwise } b_{ij1} = d_{ij1} \wedge b_{ij2} = d_{ij2} \text{ for all other dyads } ij \right\}. \end{aligned} \tag{16}$$

Next consider the conditional likelihood of observing $\mathbf{D}^3 = \mathbf{d}^3$ given that $\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)$:

$$\ell^{css}(\mathbf{d}^3; \theta, \mathbf{A}, \pi) = \frac{\Pr(\mathbf{D}^3 = \mathbf{d}^3; \theta, \mathbf{A}, \pi)}{\sum_{\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)} \Pr(\mathbf{D}^3 = \mathbf{b}^3; \theta, \mathbf{A}, \pi)}. \tag{17}$$

The ‘css’ superscript denotes “conditional star-system” (likelihood). Like, (13) introduced earlier, (17) is invariant to the population values of $\mathbf{A}$ and π (see below). Hence we will often just write $\ell^{css}(\mathbf{d}^3; \theta) = \ell^{css}(\mathbf{d}^3; \theta, \mathbf{A}, \pi)$ in order to economize on notation.

Define:

$$b(q_{ij}; \theta) \stackrel{def}{=} \exp\left(-\alpha (d_{ij2} - d_{ij1})(d_{ij3} - d_{ij0}) - \beta \left([d_{ij3} - d_{ij1}](r_{ij2} - r_{ij0}) - [d_{ij3} - d_{ij2}](r_{ij1} - r_{ij0})\right)\right), \quad (18)$$

with $q_{ij} \stackrel{def}{=} (d_{ij0}, d_{ij1}, d_{ij2}, d_{ij3}, r_{ij0}, r_{ij1}, r_{ij2})'$.

Let $\mathcal{G}_p = \{\mathcal{S}_{p1}, \dots, \mathcal{S}_{p\mathbf{m}_{p,N}}\}$ denote the set of $\mathbf{m}_{p,N}$ identifying p -star systems contained in $\mathbf{D}^3 = \mathbf{d}^3$ for $p = 2, \dots, K$. We use the bold subscripts $\mathbf{i} = \mathbf{1}, \dots, \mathbf{m}_{p,N}$ to index the identifying star systems in $\mathcal{G}_p$ in order to emphasize that they’re composed of *multiple* ($p \geq 2$) agents.

Let $\mathcal{S}_p$ denote a generic identifying system in $\mathcal{G}_p$. Next let $\mathbf{m}_{2:K,N} \stackrel{def}{=} \sum_{p=2}^K \mathbf{m}_{p,N}$ denote the *total* number of identifying star systems of size K or less. Finally, for a p -star system $\mathcal{S}_{pi} \in \mathcal{G}_p$, define the dyad set $\bar{\mathcal{S}}_{pi} = \mathcal{S}_{pi} \cup \mathcal{S}_{pi}^{fd} \cup \mathcal{S}_{pi,o}^{pd}$ with $\mathcal{S}_{pi}^{fd} = \{ij \in \mathcal{D}_p^{fd} | (i, j) \in v(\mathcal{S}_{pi}) \times v(\mathcal{S}_{pi})\}$ and $\mathcal{S}_{pi,o}^{pd} = \{ij \in \mathcal{D}_{p,o}^{pd} | (i, j) \in v(\mathcal{S}_{pi}) \times \mathcal{V}^o\}$.

Our main result about $\ell^{css}(\mathbf{d}^3; \theta, \mathbf{A}, \pi)$ is stated in the following theorem.

Theorem 1. (CONDITIONAL STAR SYSTEM LIKELIHOOD) *Under the dynamic network formation process defined by equations (3), (4), (5) and logistic random utility shocks, the conditional likelihood of the event $\mathbf{D}^3 = \mathbf{d}^3$ given $\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)$, equation (17) above, exhibits the equivalent representation:*

$$\begin{aligned} \ell^{css}(\mathbf{d}^3; \theta, \mathbf{A}, \pi) &= \prod_{p=2}^K \prod_{\mathbf{i}=1}^{\mathbf{m}_{p,N}} \left[1 + \prod_{ij \in \bar{\mathcal{S}}_{pi}} b(q_{ij}; \theta) \right]^{-1} \\ &= \prod_{p=2}^K \prod_{\mathbf{i}=1}^{\mathbf{m}_{p,N}} \frac{\exp(t'_{pi}\theta)}{1 + \exp(t'_{pi}\theta)} \end{aligned} \quad (19)$$

where $t_{pi} = (t_{1,pi}, t_{2,pi})'$, with

$$\begin{aligned} t_{1,pi} &= \sum_{ij \in \bar{\mathcal{S}}_{pi}} (d_{ij2} - d_{ij1})(d_{ij3} - d_{ij0}) \\ t_{2,pi} &= \sum_{ij \in \bar{\mathcal{S}}_{pi}} \left([d_{ij3} - d_{ij1}](r_{ij2} - r_{ij0}) - [d_{ij3} - d_{ij2}](r_{ij1} - r_{ij0}) \right). \end{aligned}$$

Proof. See Appendix B.

As in analogous results for single-agent dynamic binary choice models (e.g., Cox, 1958; Chamberlain, 1985; Honoré and Kyriazidou, 2000), Theorem 1 codifies a model implication invariant to the value of $\mathbf{A}$; allowing for fixed effects identification of $\theta_0 = (\alpha_0, \beta_0)'$. The result has a number of practical and theoretical implications.

First, direct evaluation of (17) requires enumeration of the set $\mathbb{B}_K(\mathbf{d}^3)$ (with cardinality $2^{\sum_{p=2}^K \mathbf{m}_{p,N}}$); the so called “intractable denominator problem”. Equation (19) shows that we can instead evaluate the log-likelihood by computing a sum of just $\sum_{p=2}^K \mathbf{m}_{p,N}$ elements. This is a substantial simplification.

Second, the product structure of (19) hints at the *possibility* of some form of (conditional) independence across its multiplicands. Of course, such independence is not inevitable,¹¹ but we show in Section 3 that the multiplicands in (19) are indeed conditionally independent of one another. We utilize this structure to derive single network, many agent, sampling distribution approximations.

Third, observe that three types of dyads make nontrivial contributions to (19): (i) those dyads which directly make up identifying star systems, $\bigcup_{p=2}^K \mathcal{D}_p$; (ii) those dyads where both agents belong to a star system, but the dyad itself does not, $\bigcup_{p=2}^K \mathcal{D}_p^{fd}$; and, finally, (iii) all those dyads with one agent belonging to a star system and one agent not, $\bigcup_{p=2}^K \mathcal{D}_{p,o}^{pd}$. This highlights how the entire architecture of the neighborhood surrounding a star-system contributes to identifying θ .

A proof of Theorem 1 is available in Appendix B. It involves the application of Lemmas 1 to 3 to navigate from (17) to (19). The proof provides insight into the nature of both our problem and proposed solution. It also highlights connections between our analysis and prior work on dynamic discrete choice panel data. For example, we use several “tricks” introduced by Honoré and Kyriazidou (2000) in our proof.

In the single-agent setting, Cox (1958) compares the *relative* frequencies of the link sequences $d_{ij0}01d_{ij3}$ and $d_{ij0}10d_{ij3}$ to infer the strength of state-dependence in the presence of unobserved heterogeneity.¹² *Movers* – agents who revise their choice between periods $t = 1$ and $t = 2$ – contribute to identifying state-dependence.

In the multi-agent network setting, the *relative* frequencies of the link sequences $d_{ij0}01d_{ij3}$ and $d_{ij0}10d_{ij3}$ similarly provides information about the signs and magnitudes of α_0 and β_0 . However, we must confine analysis to sets of dyads – *star systems* in our nomenclature – that,

¹¹Indeed composite likelihoods with dependence across multiplicands, feature widely in the econometrics literature on networks; see, for example, Graham (2020).

¹²The notation “ $d_{ij0}01d_{ij3}$ ” is for the link sequence $d_{ij0}, d_{ij1} = 0, d_{ij2} = 1$ and d_{ij3} .

in addition to changing their link status across periods $t = 1$ and $t = 2$, are also embedded in stable neighborhoods (see Definition 4). We separately learn about α_0 versus β_0 from star systems embedded in neighborhoods with varying types of local network architecture. The need to condition on neighborhood stability arises because of interdependencies in link decisions across dyads.

Consider two network sequences, *identical every respect*, except that in the first one dyad ij 's link history is $d_{ij0}01d_{ij3}$, while in the second it is $d_{ij0}10d_{ij3}$. The second network sequence can be derived by permuting the period $t = 1$ and $t = 2$ link decisions of *just a single dyad*. If linking decisions were conditionally independent across dyads, then the likelihoods associated with these two network sequences would differ by only a single term (corresponding to the direct likelihood contribution of dyad ij). When linking decisions are interdependent, however, these two likelihoods generally have many terms different, even though the two network sequences are nearly identical. That two nearly identical network sequences may have very different likelihoods is a consequence of the interdependence in linking decisions across dyads induced by a structural taste for transitivity.

To see this, consider the effect, on the form of the likelihood, of changing d_{ij1} from zero to one. The effect of such a small change on the structure of the likelihood is complicated. First, due to state dependence, this change alters the incentive for i and j to form a link in period $t = 2$. Second, the period $t = 2$ incentives for other agents to link with either i or j may change. This occurs if r_{il1} changes when d_{ij1} does, as would occur if l and j are linked in period $t = 1$. In that case, the presence of a period $t = 1$ link between i and j creates an opportunity for i and l to engineer triadic closure in period $t = 2$ by linking. Introducing a ij link in period $t = 1$ therefore increases the incentives for certain other links to form in period $t = 2$.¹³

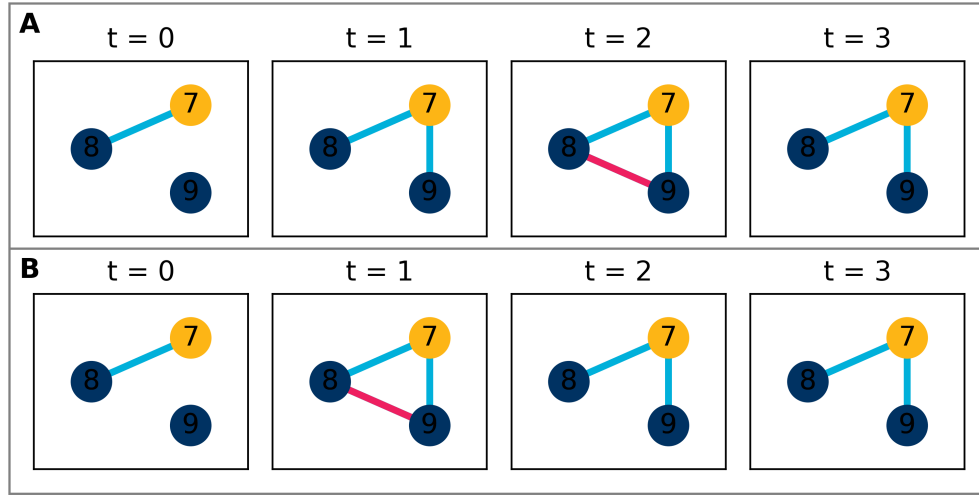
An implication of Theorem 1 is that we can control these cascading effects on the likelihood by restricting how the neighborhood surrounding a star system, say $\mathcal{S}_{pi}$, evolves. A simultaneous swapping of d_{ij1} and d_{ij2} for all dyads $ij \in \mathcal{S}_{pi}$ leaves the *net* likelihood contribution of all dyads outside of the set $\bar{\mathcal{S}}_{pi}$ unchanged. Only dyads' likelihood contributions where at least one agent is in $v(\mathcal{S}_{pi})$ are affected by swapping the edges in $\mathcal{S}_{pi}$. Specifically, while the likelihood contributions of dyads outside the set $\bar{\mathcal{S}}_{pi}$ may nominally differ, it turns out that, after permuting terms, they can be shown to be identical (this is shown carefully in the Proof). In contrast, the contributions of those dyads in $\bar{\mathcal{S}}_{pi}$ substantively *do* differ after swapping d_{ij1} and d_{ij2} . Crucially, however, the effect of $\mathbf{A}$ on the likelihood does *not* vary across such swaps. This is why $\mathbf{A}$ does not appear in (19).

¹³The change in d_{ij1} does not affect period two link incentives for pairs that do not include either i or j . This is because a change in d_{ij1} does not alter r_{kl1} for such pairs.

Examples of identifying star systems

To understand the implications of Theorem 1 it is helpful to consider some explicit examples of identifying star systems. These examples provide insight into how “swapping the edges” of star systems generate a reference set of networks which conditions away the influence of $\mathbf{A}$, while retaining information about θ .

Figure 2: Identifying a taste for transitive ties



Source: Authors' calculations.

Notes: The top (Panel A) and bottom (Panel B) rows depict two network sequences. In the top row agents 8 and 9 link in period 2, but not in period 1 (Rose Garden colored edge). In the bottom row they link in period 1, but not in period 2. Observe (i) agents 8 and 9 constitute a stable two star system (since $d_{781} = d_{782}$ and $d_{791} = d_{792}$) and (ii) while they share a link in common (with agent 7) in period $t = 1$, they do not in period $t = 0$. Consequently the payoff 8 and 9 get from forming a link is higher in period $t = 2$ than in period $t = 1$. In period 2 the link generates utility from engineering ‘triadic closure’, no such utility gain is generated by a period 1 link. Therefore, the top network sequence arises more *relatively* frequently than the bottom in the presence of a structural taste for transitivity in links.

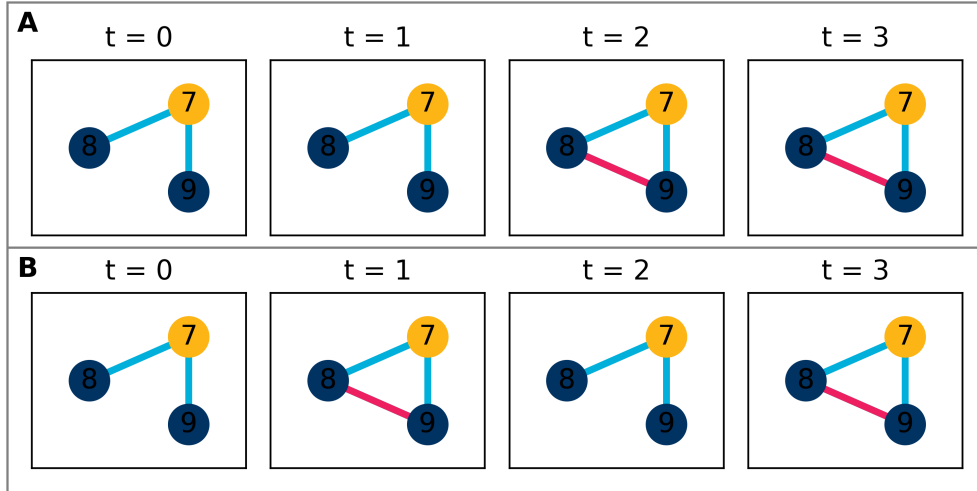
Our examples are shown in Figures 2 through 5. In the figures those nodes which belong to an identifying star system are colored “Berkeley Blue”, while neighboring nodes are colored “California Gold”. Star system edges are colored “Rose Garden”. Each row, or panel, of the figures shows a different star system edge swap.

Our first example is shown in Figure 2. Agents 8 and 9 (the two Berkeley Blue nodes) constitute an identifying 2 star system. Further observe that, while these two agents share agent 7 (the California Gold node) as a common friend in period $t = 1$, they do not in period

$t = 0$. Therefore, their returns to linking in period $t = 2$, where they can reap the benefits of triadic closure, are higher than the corresponding returns in period $t = 1$. In the presence of a structural taste for transitivity, $\beta > 0$, we will observe the top sequence *relatively* more frequently than the bottom sequence.

Figure 3 develops an example of how the relative frequency of two different network sequences provides information about the state dependence parameter α . Here the intuition parallels that familiar from the single agent binary choice case (e.g., Cox, 1958; Heckman, 1978; Chamberlain, 1985). Namely ‘runs’ in a binary choice sequence, holding the overall number of 1’s fixed, suggests true state dependence. In our setting, identification additionally requires holding the payoffs from triadic closure fixed. See the notes to the figure for more details.

Figure 3: Identifying a structural state dependence



Source: Authors’ calculations.

Notes: See the notes to Figure 2. In this example $r_{890} = r_{891} = 1$, so agents 8 and 9 accrue the same transitivity payoff whether linking in periods 1 or 2. However, $d_{890} = 0$ and $d_{893} = 1$, suggesting that, in the presence of state dependence ($\alpha > 0$), the top sequence (Panel A) will occur *relatively* more frequently than the bottom (Panel B). The intuition in this case is very similar to that underlying the results of Cox (1958), Heckman (1978) and Chamberlain (1985).

Figure 4 presents a more complicated example. In this figure dyads 12 and 13 form an identifying 3 star system with 1 as the center agent. The neighborhood of this system includes agents 4, 5, 6 and 7 (colored in California Gold). Although tedious, it is straightforward to check that the stability requirements of Definition 4 are satisfied. Inspecting the top sequence, we see that agents 1 and 2 share three links in common in period $t = 0$ (with agents 4, 5 and

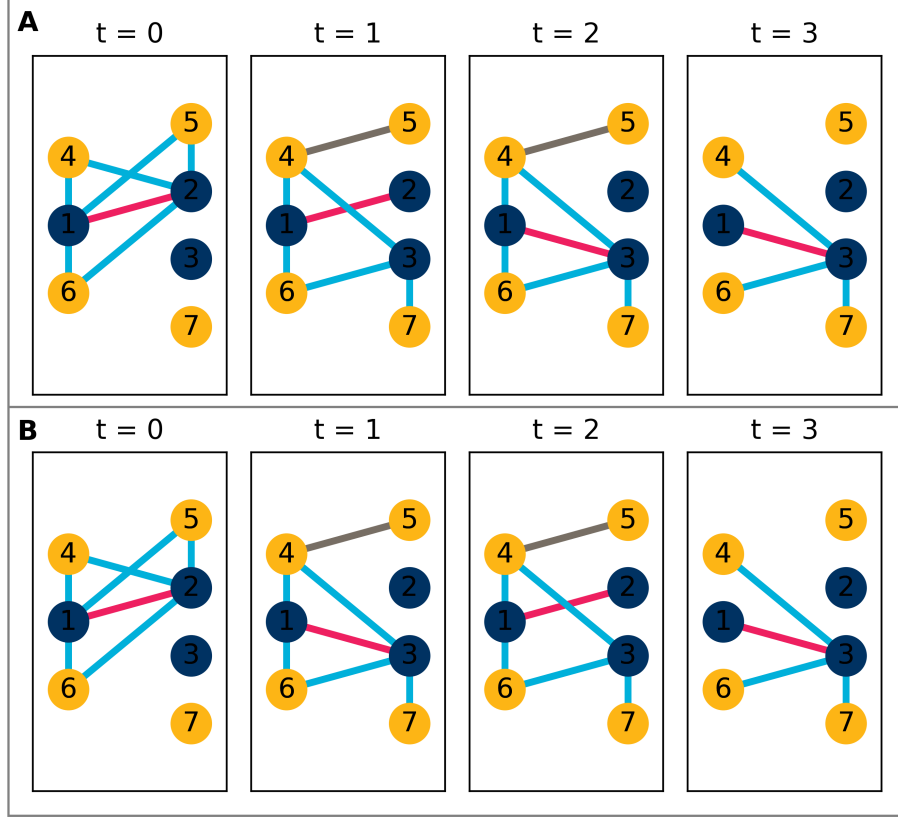
6). Whereas they share no links in common in subsequent periods. In contrast, agents 1 and 3 share no connections in common in period $t = 0$, but they do share two connections (with agents 4 and 6) in periods $t = 1$ and $t = 2$. This neighborhood architecture implies that the payoff from transitive closure will be larger for agents 1 and 2 if they connect in period $t = 1$ versus period $t = 2$ (since they share many friends in common in period $t = 0$). In contrast, the transitivity-driven gains for agents 1 and 3 are greater in period $t = 2$ versus period $t = 1$ (since these two agents shared several connections in common in period $t = 1$, but not in period $t = 0$). Therefore, in the presence of a structural taste for transitivity, we will observe the top sequence *relatively* more frequently than the bottom one. The reader can verify that the top sequence is also more likely in the presence of structural state dependence.

Our final example, shown in Figure 5, combines the star systems shown in Figures 2 and 4 into a single network. Note that the agents forming the two different star systems – $\mathcal{S}_3 = \{12, 13\}$ and $\mathcal{S}_2 = \{89\}$ – although not directly connected, have overlapping neighborhoods (both star systems include agents connected to agent 7). See Lemmas 1 and 2 above.

Because there are two identifying star systems, the set $\mathbb{B}_K(\mathbf{d}^3)$ includes four elements (based on whether we swap the edges, or not, in each of the two star systems). For a given value of θ , we can evaluate (19). Values of θ making the observed network sequence, say the one depicted in Panel A, relatively more likely than the other sequences (those in Panels B, C and D) are more data consonant.

Each of the four networks depicted in Figure 5 share common values for those statistics sufficient for $\mathbf{A}$ and π (see Proposition 1 above). However, the networks *do* vary in their numbers of consecutive links, $\sum_{i < j} \sum_{t=1}^3 d_{ijt-1} d_{ijt}$, and attempts to engineer triadic closure, $\sum_{i < j} \sum_{t=1}^3 r_{ijt-1} d_{ijt}$. If, among all networks in the reference set (Panels A, B, C, & D), the observed network (Panel A) has *relatively* more consecutive links and attempts to close triangles, then we can conclude that α and β are larger. The maximum likelihood estimate (MLE) chooses $\hat{\theta} = (\hat{\alpha}, \hat{\beta})'$ to make likelihood of the network shown in Panel A as large as possible, *relative*, to that of Panels B, C or D. In the next section, we present conditions under which the MLE consistently estimates θ_0 .

Figure 4: A more complicated example



Source: Authors' calculations.

Notes: The top row (Panel A) depicts an identifying 3 star, $\mathcal{S}_3 = \{12, 13\}$. In $t = 0$ dyad 12 shares friends in common with agents 4, 5 and 6. They do not share any friends in common in later periods. In contrast, dyad 13 has no friends in common in $t = 0$, but *does* share connections in common with agents 4 and 6 in $t = 1$. If $\beta > 0$, we would expect to observe the top sequence relatively more frequently than the bottom one (Panel B). This follows since it maximizes the payoffs from transitive closure.

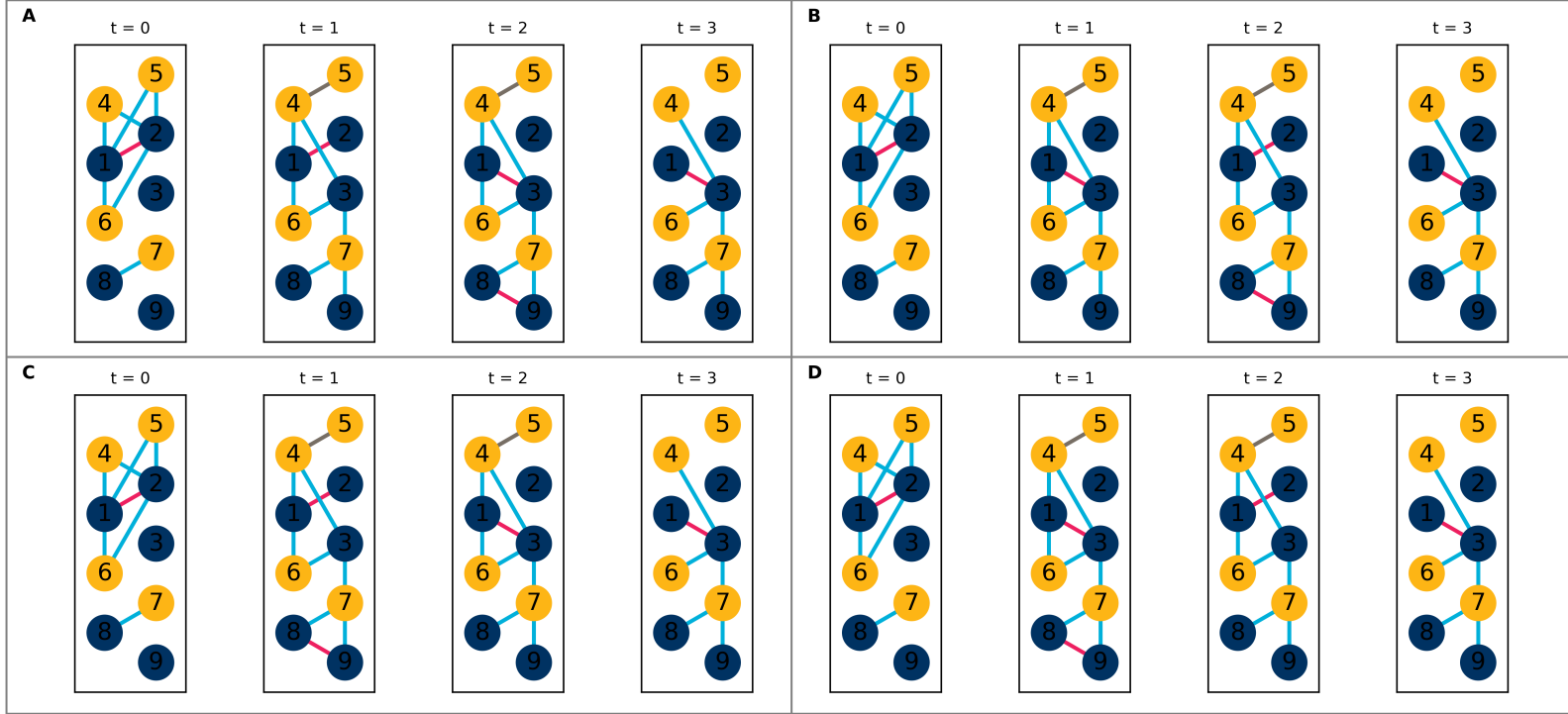
3 Inference on α and β

In this section we develop two approaches to conditional inference on $\theta = (\alpha, \beta)'$. The first method is exact (and similar) (cf., Pelican and Graham, 2026).¹⁴ This method exploits the fact that the conditional distribution of $\mathbf{D}^3$ given $\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ is a parametric family indexed by θ alone.

¹⁴By 'similar' we mean that the size of the test is invariant to the precise instance of the conditioning sufficient statistics. See Ferguson (1967) for a formal definition and/or the related discussion in Pelican and Graham (2026).

The second method involves constructing an estimate of θ by maximizing (19) and characterizing its sampling properties as the network grows large, $N \rightarrow \infty$. This asymptotic argument is delicate: we observe only a single network for $T = 3$ periods (plus the initial condition). Consistency and asymptotic normality requires the number of identifying star-systems within the network to grow with N ; a condition which places subtle restrictions on the form of population network generating process.

Figure 5: Conditional star system (css) likelihood



Source: Authors' calculations.

Notes: Panel A depicts a hypothetical observed network sequence containing two identifying star systems: the 3-star $\{1, 2, 3\}$ and the 2-star $\{6, 7\}$. In this sequence, there are 18 instances of links persisting across two adjacent periods, $\sum_{i < j} \sum_{t=1}^3 d_{ij,t-1} d_{ij,t} = 18$, and 16 attempts to engineer triadic closure, $\sum_{i < j} \sum_{t=1}^3 r_{ij,t-1} d_{ij,t} = 16$. In Panel B, the period-1 and period-2 edges of the 3-star system are swapped, reducing the persistence total to 16 and the transitivity total to 13. In Panel C, only the period-1 and period-2 edges of the 2-star system are swapped. This leaves the persistence total unchanged at 18 but reduces the transitivity total to 15. In Panel D, both star systems' period-1 and period-2 edges are swapped, yielding a persistence total of 16 and a transitivity total of 12.

Exact inference on α and β

This subsection describes an *exact* method of inference on θ . The basic approach is due to Fisher (1922). The main difficulty, in the present context, involves enumerating the set $\mathbb{B}_K(\mathbf{d}^3)$; a problem which we solve by simulation (as do McDonald et al. (2007) and Pelican and Graham (2026) in related contexts).

A Metropolis-Hastings algorithm

Our main tool is Algorithm 1, which describes a Metropolis-Hastings procedure for constructing a random draw from the conditional distribution of $\mathbf{D}^3$ given that $\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)$. This distribution, for a null structural parameter of $\theta = \theta_0$, places a probability weight of $\ell^{css}(\mathbf{b}^3; \theta_0)$ on each $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ (see equation (19) of Theorem 1 above).

Algorithm 1 stochastically moves between the elements of $\mathbb{B}_K(\mathbf{d}^3)$ by (at each iteration $s = 1, \dots, \sigma$) randomly selecting a stable star system and “swapping its edges” (i.e., permuting its period 1 and 2 edges). More precisely, at each iteration we (i) stay at the current state with probability $0 < \rho < 1$ or (ii), with probability $1 - \rho$, pick a stable star system at random and swap its edges. In case (ii) the swap causes the algorithm to move from its current state, say $\mathbf{B}^{3,(s)}$, to a *proposed* new state – also in $\mathbb{B}_K(\mathbf{d}^3)$ by construction – of $\tilde{\mathbf{B}}^{3,(s+1)}$. This proposal is then accepted according to the rule

$$\mathbf{B}^{3,(s+1)} = \begin{cases} \tilde{\mathbf{B}}^{3,(s+1)} & \text{if } -\theta'_0 w(\mathbf{B}^{3,(s)}, \tilde{\mathbf{B}}^{3,(s+1)}) \leq V^{(s)} \\ \mathbf{B}^{3,(s)} & \text{otherwise} \end{cases},$$

with $V^{(s)} \sim \text{Exponential}(1)$ ¹⁵ and

$$w(\mathbf{B}^{3,(s)}, \tilde{\mathbf{B}}^{3,(s+1)}) = \left(\frac{\sum_{i < j} \sum_{t=1}^3 \left[\tilde{B}_{ijt-1}^{(s+1)} \tilde{B}_{ijt}^{(s+1)} - B_{ijt-1}^{(s)} B_{ijt}^{(s)} \right]}{\sum_{i < j} \sum_{t=1}^3 \left[\tilde{S}_{ijt-1}^{(s+1)} \tilde{B}_{ijt}^{(s+1)} - S_{ijt-1}^{(s)} B_{ijt}^{(s)} \right]} \right).$$

Recall that $S_{ijt} = \sum_k B_{ikt} B_{jkt}$ and define $\tilde{S}_{ijt-1}$ analogously. Note that $\theta'_0 w(\mathbf{B}^{3,(s)}, \tilde{\mathbf{B}}^{3,(s+1)})$ equals the logarithm of the likelihood ratio $\ell^{css}(\tilde{\mathbf{B}}^{3,(s+1)}; \theta_0) / \ell^{css}(\mathbf{B}^{3,(s)}; \theta_0)$. Inspection of the expressions above reveal, assuming (for clarity of exposition) that the elements of θ_0 are positive, that Algorithm 1 accepts any move to a network sequence with (sufficiently) “more” link persistence and/or transitivity. However, moves with reduced link persistence and/or transitivity also occur with positive probability. Because the conditional star system likelihood assigns higher probability to network sequences with more link persistences and/or

¹⁵Recall that if $U^{(s)} \sim \mathcal{U}[0, 1]$, then $-\ln U^{(s)} \sim \text{Exponential}(1)$.

Algorithm 1 MARKOV DRAW ALGORITHM

Inputs: A sequence of adjacency matrices $(\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) = \mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ with a set $\mathcal{G}$ of $\mathbf{m}_{2:K,N}$ stable star systems of size $p = 2, \dots, K$; a null value for the common parameter, θ_0 ; a mixing time σ .

Procedure:

1. Set $s = 0$.
2. With probability $1 - \rho$ go to step 3, with probability ρ go to step 4.
3. Draw a stable star system $\mathcal{S}$ uniformly at random from $\mathcal{G}$ and permute its period 1 and 2 edges (as currently configured in $\mathbf{b}^3$) generating $\tilde{\mathbf{b}}^3 \in \mathcal{N}(\mathbf{b}^3) \subseteq \mathbb{B}_K(\mathbf{d}^3)$.
 - (a) Draw $U \sim \text{Uniform}[0, 1]$.
 - (b) **If** $\min \left\{ 1, \frac{\ell^{css}(\tilde{\mathbf{b}}^3; \theta_0)}{\ell^{css}(\mathbf{b}^3; \theta_0)} \right\} \geq U$.
 - i. Set $\mathbf{b}^3 = \tilde{\mathbf{b}}^3$
 - ii. Go to step 4.
 - (c) **Else** go to step 4.
4. Set $s = s + 1$.
 - (a) **If** $s = \sigma$ return $\mathbf{b}^3$.
 - (b) **Else** go to step 2.

Output: A random draw, $\mathbf{b}^3$, from $\ell^{css}(\cdot; \theta_0)$.

transitivity, the Markov chain defined by Algorithm 1 spends more time in regions in $\mathbb{B}_K(\mathbf{d}^3)$ assigned high probability by the null model; but it also visits all network sequences in this reference set with positive probability.

Let $\mathbf{B}^{3,(0)} = \mathbf{d}^3$ – the observed network sequence – initialize the chain and consider the simulated chain $\mathbf{B}^{3,(1)}, \mathbf{B}^{3,(2)}, \dots, \mathbf{B}^{3,(\sigma)}$. For σ large enough, it is straightforward to verify¹⁶ that

$$\mathbf{B}^{3,(S)} \sim \ell^{css}(\cdot; \theta_0).$$

That is, the ergodic distribution of the Markov process described above is such that it visits each network in $\mathbb{B}_K(\mathbf{d}^3)$ with a long-run frequency of

$$\Pr(\mathbf{B}^{3,(s)} = \mathbf{b}^3 | \mathbf{B}^{3,(s)} \in \mathbb{B}_K(\mathbf{d}^3); \theta_0) = \ell^{css}(\mathbf{b}^3; \theta_0)$$

for $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ and with zero frequency for $\mathbf{b}^3 \notin \mathbb{B}_K(\mathbf{d}^3)$.

Theorem 2. (NULL SAMPLER) *The output of Algorithm 1, for σ large enough, is a random draw from $\ell^{css}(\cdot; \theta_0)$.*

A proof of Theorem 2 can be found in Appendix B.3. The ‘large enough’ proviso is a limitation. However, based upon related work, we expect the mixing properties of this MCMC chain to be quite good. A rule-of-thumb is to run the chain until the edges of each identifying star system are swapped at least two times, viewing further draws as from the target stationary distribution.

Let $(\mathbf{A}, \pi) \in \Delta$ denote the nuisance parameter space and $\gamma = (\theta, \mathbf{A}, \pi)$ the full model parameter. Define the set $\Gamma_0 = \{(\theta, \mathbf{A}, \pi) : \theta = \theta_0, (\mathbf{A}, \pi) \in \Delta\}$. Our goal is to construct a confidence set by inverting a two-sided test of the composite null

$$H_0 : \gamma \in \Gamma_0. \tag{20}$$

Under (20) the structural parameter is fixed at $\theta = \theta_0$, while the nuisance parameter $(\mathbf{A}, \pi)$ is allowed to range freely over Δ ; hence the ‘composite null’ nomenclature. See Pelican and Graham (2026) for more discussion in a related context.

Let $v(\mathbf{d}^3)$ be some test statistic. A starting point might be a count of the number of links persisting for two periods *plus* the number of two-stars ‘closing’ into triangles the following

¹⁶The Markov chain is finite, irreducible and aperiodic and hence ergodic. The claimed stationary distribution can be derived using, for example, Theorem 7.10 of Mitzenmacher and Upfal (2005)

period:

$$v(\mathbf{d}^3) = \sum_{i < j} \sum_{t=1}^3 d_{ijt-1} d_{ijt} + \sum_{i < j} \sum_{t=1}^3 r_{ijt-1} d_{ijt}.$$

Under H_0 the conditional distribution of $v(\mathbf{D}^3)$ given $\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ equals

$$F_{v(\mathbf{D}^3)|\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)}(c|H_0) = \sum_{\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)} \mathbf{1}(v(\mathbf{b}^3) \leq c) \cdot \ell^{css}(\mathbf{b}^3; \theta_0, \mathbf{A}, \pi),$$

which can be approximated by

$$\hat{F}_{v(\mathbf{D}^3)|\mathbf{D}^3 \in \mathbb{B}_K(\mathbf{d}^3)}(c|H_0) = \frac{1}{S} \sum_{s=1}^S \mathbf{1}(v(\mathbf{B}^{3,(s)}) \leq c), \quad (21)$$

where $s = 1, \dots, S$ indexes draws from the ergodic distribution of the Markov chain defined by Algorithm 1. If, for a given $\theta = \theta_0$, $v(\mathbf{D}^3)$ lies between the $\alpha/2$ and $1 - \alpha/2$ quantiles of distribution (21), then we can conclude that θ_0 lies within a $1 - \alpha$ confidence set for θ . By repeating this process for an appropriately chosen grid of θ_0 values we can compute the boundaries of this confidence set.

Observe that the coverage of this confidence set *exactly equals* $1 - \alpha$ (modulo simulation/numerical approximation error which can be minimized with sufficient compute).¹⁷ Invariance of the conditional distribution, to where $(\mathbf{A}, \pi)$ in Δ lies, ensures that the confidence set is *similar* (in the sense that its coverage is exactly $1 - \alpha$ irrespective of the population value of the nuisance parameter). We leave a more careful exploration of exact inference on θ_0 to future work.

Asymptotic inference on α and β

In this section we study the properties of the maximizer, $\hat{\theta}$, of the conditional star system log-likelihood (see Theorem 1 above). Specifically, we consider the approximate repeated sampling properties of $\hat{\theta}$. The econometrician observes a single network sequence, $\mathbf{D}^3 = \mathbf{d}^3$. We consider the behavior of $\hat{\theta}$ across repeated draws of $\mathbf{D}^3$. Our asymptotic approximation is with respect to the number of agents in these single network sequence draws (i.e., $N \rightarrow \infty$).

¹⁷There is also a small amount of coverage distortion induced by the fact that $v(\mathbf{D}^3)$ is a discrete random variable with finite support. This can be formally resolved by inverting a randomized test. However, this issue is practically irrelevant when N is modestly large.

Notation and setup

In what follows, it will be convenient to iterate over all star systems, whether identifying or not. We use $\mathbf{i}$ to index a generic star system (of order $p = 2, \dots, K$) and, also, in an abuse of notation the dyad set which defines it. So that $\mathbf{i}$ can denote both a counting number and a set of the form $\mathbf{i} = \{i_1 i_2, i_1 i_3, \dots, i_1 i_p\}$. We further use $v(\mathbf{i})$ to denote the vertex set $\{i_1, \dots, i_p\}$.

Observe that there are $\mathbf{n}_p \stackrel{\text{def}}{=} \binom{N}{p} \times p$ star systems of size p . To keep track of whether a star system is identifying, we define the binary indicator variable Z_{pi} , which equals 1 if the $\mathbf{i}^{th}$ star system of size p is identifying and 0 otherwise. Define $\mathbf{Z} := (Z_{pi})_{2 \leq p \leq K, \mathbf{i} = 1, \dots, \mathbf{n}_p}$.

Recall that $Q_{ij} = (D_{ij0}, D_{ij1}, D_{ij2}, D_{ij3}, R_{ij0}, R_{ij1}, R_{ij2})'$. Next define the tuples

$$Z_{pi} \mathbf{Q}_{pi} := Z_{pi} \cdot \left(Q_{ij} : i \in v(\mathbf{i}), j \in \mathcal{V} - \bigcup_{q=2}^K \bigcup_{\mathbf{k} \in \{\mathbf{k}=1, \dots, \mathbf{n}_p : Z_{q\mathbf{k}}=1\}} v(\mathbf{k}) + v(\mathbf{i}) - i \right).$$

Here we multiply by Z_{pi} to denote the conditional selection (or filtering) of those Q_{ij} associated with stable star system $\mathbf{i}$. The vertex set $\mathcal{V} - \left\{ \bigcup_{q=2}^K \bigcup_{\mathbf{k} \in \{\mathbf{k}=1, \dots, \mathbf{n}_p : Z_{q\mathbf{k}}=1\}} v(\mathbf{k}) + v(\mathbf{i}) - i \right\}$ includes all agents j not in the vertex set of stable star systems *other than* $\mathbf{i}$. Hence $Z_{pi} \mathbf{Q}_{pi}$ returns all Q_{ij} where at least one of i or j is in $v(\mathbf{i})$, $\mathbf{i}$ is identifying, and every Q_{ij} appears at most once in $Z_{pi} \mathbf{Q}_{pi}$.

At the heart of our large sample theory is the following conditional independence result:

Lemma 4. (CONDITIONAL INDEPENDENCE) *The tuples $Z_{pi} \mathbf{Q}_{pi}$ are independent across indices $\mathbf{i} \in \{i_1 i_2, \dots, i_1 i_q : q \leq K\}$ conditional on $\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z} := (\mathbf{Z}_{q,\mathbf{i}})_{\mathbf{i}, q \leq K}$ and $\mathbf{A}$.*

A proof is in Appendix A.2.

Define the normalized version of the conditional star system log-likelihood:

$$\begin{aligned} \hat{Q}_N(\theta) &= \frac{1}{\mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{pi} \right]} \ln \left[\prod_{p=2}^K \prod_{\mathbf{i}=1}^{\mathbf{m}_{p,N}} \left[1 + \prod_{ij \in \mathcal{S}_{pi}} b(Q_{ij}; \theta) \right] \right]^{-1} \\ &= \frac{1}{\mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{pi} \right]} \sum_{p=2}^K \sum_{\mathbf{i}=\{i_1 i_2, \dots, i_1 i_p\}} Z_{pi} \cdot \ln \left[\frac{\exp(T'_{pi} \theta)}{1 + \exp(T'_{pi} \theta)} \right], \end{aligned} \quad (22)$$

where the second equality follows from the notation defined immediately above. Our estimate coincides with $\hat{\theta} = \arg \max_{\theta} \hat{Q}_N(\theta)$. By Lemma 4 the summands of $\hat{Q}_N(\theta)$ are conditionally independent of one another. However the number of these summands which are non-zero is a random variable; equal to the number of identifying star systems in $\mathbf{D}^3$. As long as the number of such systems, $\sum_{p,\mathbf{i}} Z_{pi}$, grows as $N \rightarrow \infty$, the conditional independence structure

established by Lemma 4 is sufficient, under some additional regularity conditions, to formally establish consistency and asymptotic normality of $\hat{\theta}$.

Our consistency result is based on the following four assumptions.

Assumption 1. (PARAMETER SPACE) $\theta_0 \in \text{int}(\Theta)$, a closed ball in $\mathbb{R}^2$.

Assumption 2. (RANK) There exists a constant $C_1 > 0$ such that for N large enough:

$$\lambda_{\min} \left(\frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \mathbb{E} \left[\sum_{p,i} Z_{pi} T_{pi} T'_{pi} \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z} \right] \right) \geq C_1.$$

Assumption 3. (BOUNDEDNESS) There exists $C_2 > 0$ such that for N large enough:

$$\frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} \cdot \mathbb{E} \left[\|T_{pi}\|^4 \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z} \right] \leq C_2.$$

Assumption 4. (IDENTIFYING EVENTS) The sequence $\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]$ diverges to $+\infty$ as $N \rightarrow \infty$.

Assumption 4, which is key, ensures that our criterion function includes a growing number of conditionally independent summands as $N \rightarrow \infty$. Assumption 1 is standard. Assumption 2 is a rank condition, while Assumption 3 restricts the first four moments of T_{pi} .

Observe that, since T_{pi} is a direct function of $\mathbf{D}^3$, Assumptions 2 to Assumption 4 involve implicit restrictions on the distribution of $\mathbf{A}$ and the initial network condition π . For this reason, they are not straightforward to verify. Below we develop examples of data-generating processes where Assumptions 4 provably holds, and provably fails to hold. These examples gives some insight into what types of networks are likely to include a sufficient number of identifying star systems to support estimation and inference.

Of course, for any particular network under actual consideration, the researcher necessarily counts the number of identifying star systems available and directly assesses whether they arise in sufficient numbers to justify estimation (and large sample inference). Since θ only includes two elements, and the conditional star system log-likelihood is globally concave, we expect our large sample theory for $\hat{\theta}$ to be quite accurate even if settings with relatively modest numbers of identifying star systems. This accords with the limited Monte Carlo evidence summarized below.

Theorem 3. (CONSISTENCY) Under Assumptions 1 to 4, $\hat{\theta} \xrightarrow{p} \theta_0$.

A proof is available in Appendix B.

Let $\Lambda(x) \stackrel{\text{def}}{=} \frac{\exp(x)}{1+\exp(x)}$. Our demonstration of asymptotic normality requires the additional assumption:

Assumption 5. (HESSIAN) *The normalized (negative) Hessian*

$$\frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \mathbb{E} \left[\sum_{p,i} Z_{pi} \cdot \Lambda(T'_{pi}\theta_0)(1 - \Lambda(T'_{pi}\theta_0)) T_{pi} T'_{pi} \right]$$

converges to some positive definite matrix $\mathcal{I}(\theta_0)$.

Theorem 4. (ASYMPTOTIC NORMALITY) *Under Assumptions 1 to 5:*

$$\sqrt{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \cdot (\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0)).$$

A complete proof is available in Appendix B. As expected, the rate-of-convergence depends on the (expected) number of identifying star systems; corresponding to the “effective sample size” in our setting. The examples we develop in the next subsection indicate that the behavior of $\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]$ as $N \rightarrow \infty$ varies with the distribution of $\mathbf{A}$ as well as the initial network condition, π . Hence, while our main identification result, Theorem 1 above, holds for *any* $\mathbf{A}$ and π , the precision with which we can learn θ is not invariant to these nuisance parameters. Indeed, for some distributions of heterogeneity, consistent estimation is not possible. By Corollary 1, the number of identifying star systems is, at most, of order N , suggesting a maximal rate-of-convergence of $\sqrt{N}$.

Single network estimation: examples

In this subsection, by means of two examples, we explore conditions under which the number of identifying systems diverges as N grows large (Assumption 4). Such divergence is required for consistency and asymptotic normality. Whether Assumption 4 holds depends on the model’s parameters, the distribution of the fixed effects, and the initial network.

We study two models. In the first, no identifying star systems are observed (on average) when N is large. In the second, the expected number of identifying star systems grows linearly with N (the maximal rate). These two examples provide some indication of the positive possibilities associated with our method, as well as its limits.

Example 1. (DENSE INITIAL GRAPH AND COMPACTLY SUPPORTED FIXED EFFECTS) Suppose:

- the fixed effects A_{ij} have bounded support, say $|A_{ij}| \leq \bar{A}$ for all i, j almost surely.
- D_0 is an Erdos-Renyi random graph with edge probability $p \in (0, 1)$.

Under this DGP, for any i and j , $R_{ij0} \sim (N-2)p^2$. When $\beta_0 > 0$, this implies that with high probability, $-|\alpha_0| + \beta_0 R_{ij0} - \bar{A} \gg U_{ij1}$, such that $\mathbf{D}_1$ approaches a complete graph. Consequently, $\mathbf{D}_2$ and $\mathbf{D}_3$ also become “nearly complete”. In such dense graph sequences, very few edges transition between periods 1 and 2, making condition (ii) of Definition 5 difficult to satisfy.

Conversely, when $\beta_0 < 0$, we show that with high probability, $\mathbf{D}_1 = 0$. Therefore, conditional on the event that $\mathbf{D}_1 = 0$, for a given star system to be identifying it must be that: *i)* the nodes involved in the star become connected with the central node in the second period, *ii)* these nodes remain disconnected with the rest of the network and *iii)* the non central nodes remain disconnected from each other. We show that *i)*, *ii)* and *iii)* only happen with a vanishing probability, which shows that identifying stars form with small probability. Specifically, Proposition 2 demonstrates that $\sum_K \mathbf{m}_{K,N}$, the total number of identifying stars, vanishes as N grows. The complete proof is provided in Appendix B.6.

Proposition 2. *Under the DGP of Example 1:*

$$E \left(\sum_K \mathbf{m}_{K,N} \right) \rightarrow_{N \rightarrow \infty} 0.$$

Example 2. (RANDOM GEOMETRIC GRAPH) Let $z_1, \dots, z_N$ be i.i.d. uniform on the square $[0, \sqrt{N}]^2$. Fix a finite real number a and define

$$A_{ij} := \begin{cases} a, & \|z_i - z_j\| \leq r, \\ -\infty, & \|z_i - z_j\| > r. \end{cases}$$

For $D_{ij0} := \mathbf{1}(A_{ij} - U_{ij0} \geq 0)$ for all i, j , we can show that – on average – the number of stable systems of any fixed size grows linearly in the sample size N .¹⁸

Proposition 3. *Let $\mathbf{m}_{K,N}$ be the number of identifying K -systems, then¹⁹*

1. $\mathbb{E} \left[\sum_{K=2}^{N-1} \mathbf{m}_{K,N} \right] \asymp N$
2. $\mathbb{E} \left[\sum_{K=2}^{\bar{K}} \mathbf{m}_{K,N} \right] \asymp N$ for any fixed $\bar{K} \leq N-1$.

¹⁸In fact, any $\mathbf{D}_0$ such that $\mathbb{P}(D_{ij,0} = 0|A)$ is bounded away from 0 works.

¹⁹For any sequences of real numbers u_n and v_n , we write $u_n \asymp v_n$ when $u_n = O(v_n)$ and $v_n = O(u_n)$.

A formal proof is in Appendix B.7. The proof strategy relies on establishing upper and lower bounds on the probability that any specific group of nodes forms an identifying star. For the upper bound, we note that for the “spokes edges” of a star to switch between $t = 1$ and $t = 2$, all “leaf nodes” must randomly fall within a small, fixed distance, from the central node: an event whose probability shrinks rapidly with N . For the lower bound, the proof constructs an “isolated bubble” scenario: the star’s nodes are clustered tightly together, all other nodes are forced far enough away to prevent any outside interference, and the internal links follow a specific sequence of forming and breaking. Together, the two bounds imply that the probability of a specific K -star acting as an identifying star decays proportionally to $N^{-(K-1)}$. Because the total number of possible node combinations to form a K -star grows proportionally to N^K , multiplying these factors shows that the expected total number of identifying stars scales linearly with the overall network size, N . Note that by Corollary 1, this is the fastest rate the number of stable systems could reach under any DGP.

4 Monte Carlo experiments

The Monte Carlo design uses a random geometric graph to construct an “opportunity graph” for link formation (as in the second example considered in the previous section). Specifically agents are scattered uniformly on the two-dimensional plane

$$\left[0, \sqrt{N}\right] \times \left[0, \sqrt{N}\right].$$

The initial network is then generated according to the rule

$$D_{ij0} = \mathbf{1}(A_{ij} - U_{ij0} \geq 0),$$

with U_{ij0} logistic and A_{ij} taking one of two values. If the Euclidean distance between i and j is less than or equal to r , then $A_{ij} = \ln\left(\frac{0.75}{1-0.75}\right)$, otherwise A_{ij} equals negative infinity. This calibration means that, in the initial period, agents that are less than r apart link with probability 0.75, while those greater than r apart link with probability zero.

The expected degree of a randomly sampled agent in $t = 0$ is approximately $0.75\pi r^2$ (An exact expression for average degree, which takes into account boundary effects, can be calculated along the lines of Kostin (2010)). An implication of this set-up is that the initial condition is sparse: average degree does not increase with network size. By varying the value of r , we can manipulate the average degree, and hence connectivity, of the initial condition. We choose

values of r such that in large networks average degree in period $t = 0$ is 2, 3 or 4.²⁰ With an average degree of 2, the resulting initial condition consists of many small disconnected components. When average degree equals 3 a large giant component begins to form. Finally when average degree equals 4, more than half of agents are part of one giant component (see Table 1). It is well-known that a phase-transition occurs at an average degree of 3 in random geometric graphs (Penrose, 2003).

In period $t = 1, 2, 3$ links evolve according to rule (3) with $\alpha_0 = \beta_0 = 1$.²¹ A key feature of this design is that it generates homophilous link formation based on location (which is unobserved by the econometrician). Although there are no network effects in operation in period $t = 0$, measured transitivity of the initial network is quite high, with the clustering coefficient exceeding 0.4 in all cases (see Table 1). In subsequent periods transitivity and average degree increase as agents form additional links (on net) in response to state dependence and a taste for triadic closure. Link clustering in these designs therefore arises from both homophily and a taste for triadic closure, making them an appropriate test case for evaluating the small sample relevance of Theorem 4.

A typical sequence of networks, with $N = 100$ and average degree in the initial network set equal to 2, is depicted in Figure 6. The figure shows that identifying star systems can come in different forms and arise in both sparse and relatively dense regions. In the figure the larger Berkeley blue and brown nodes correspond to identifying 2-star and 3-star systems respectively, while the California gold nodes are their immediate neighbors.

Table 1: Basic Properties of Simulated Network Sequences

Asym. Degree	2			3			4		
Period	$(N - 1)\mathbb{E}[D_{it}]$	T	GC	$(N - 1)\mathbb{E}[D_{it}]$	T	GC	$(N - 1)\mathbb{E}[D_{ijt}]$	T	GC
$t = 0$	1.98	0.44	0.01	2.96	0.44	0.05	3.94	0.44	0.58
$t = 1$	2.41	0.58	0.01	3.68	0.58	0.08	4.97	0.58	0.83
$t = 2$	2.49	0.59	0.01	3.80	0.59	0.09	5.12	0.59	0.85
$t = 3$	2.50	0.59	0.01	3.82	0.59	0.09	5.14	0.59	0.85

Notes: The table reports period-specific network summary statistics across the $B = 1,000$ Monte Carlo simulations for each design ($N = 5,000$). See main text for other design details. The $(N - 1)\mathbb{E}[D_{ijt}]$ column gives the average degree, T the global clustering coefficient, or transitivity index, and GC the fraction of agents that are part of the largest giant component.

Table 2 summarizes the sampling properties of the CSS-MLE for α_0, β_0 across the different

²⁰As a point of reference McPherson et al. (2006) estimate the average size of adult Americans' core discussion networks (i.e., confidants with whom individuals discuss important matters) was about three in 1985 and two in 2004.

²¹Since agents greater than r apart will never link, average degree in the network is bounded above by πr^2 in periods $t = 1, 2, 3$. This corresponds to maximum average degrees of (approximately) 2.67, 4 and 5.33 across the three designs.

designs. Panel A presents results for the conditional 2-star system estimator, whereas Panel B reports results for the conditional 2 & 3-stars system estimator, which pools information from all stable star systems of sizes 2 *and* 3. The number of agents is set to $N = 5,000$ and we complete $B = 1,000$ Monte Carlo replications. For the designs considered here, with five thousand agents, the number of stable 2-star systems averages between 136 and 290 while the number of stable 3-star systems ranges between 32 and 60. This implies a ratio of stable 3-star systems to stable 2-star systems of about 20 – 24 percent.²²

Inspection of Table 2 shows that both estimators are approximately mean and median unbiased and that the associated Wald-based confidence intervals have coverage close to nominal 95 percent coverage. Furthermore, consistent with the theory, the conditional 2 & 3-star estimator (Panel B) is relatively more precise than its 2-star counterpart (Panel A); as indicated by both a smaller mean absolute error and mean standard deviation across all designs.

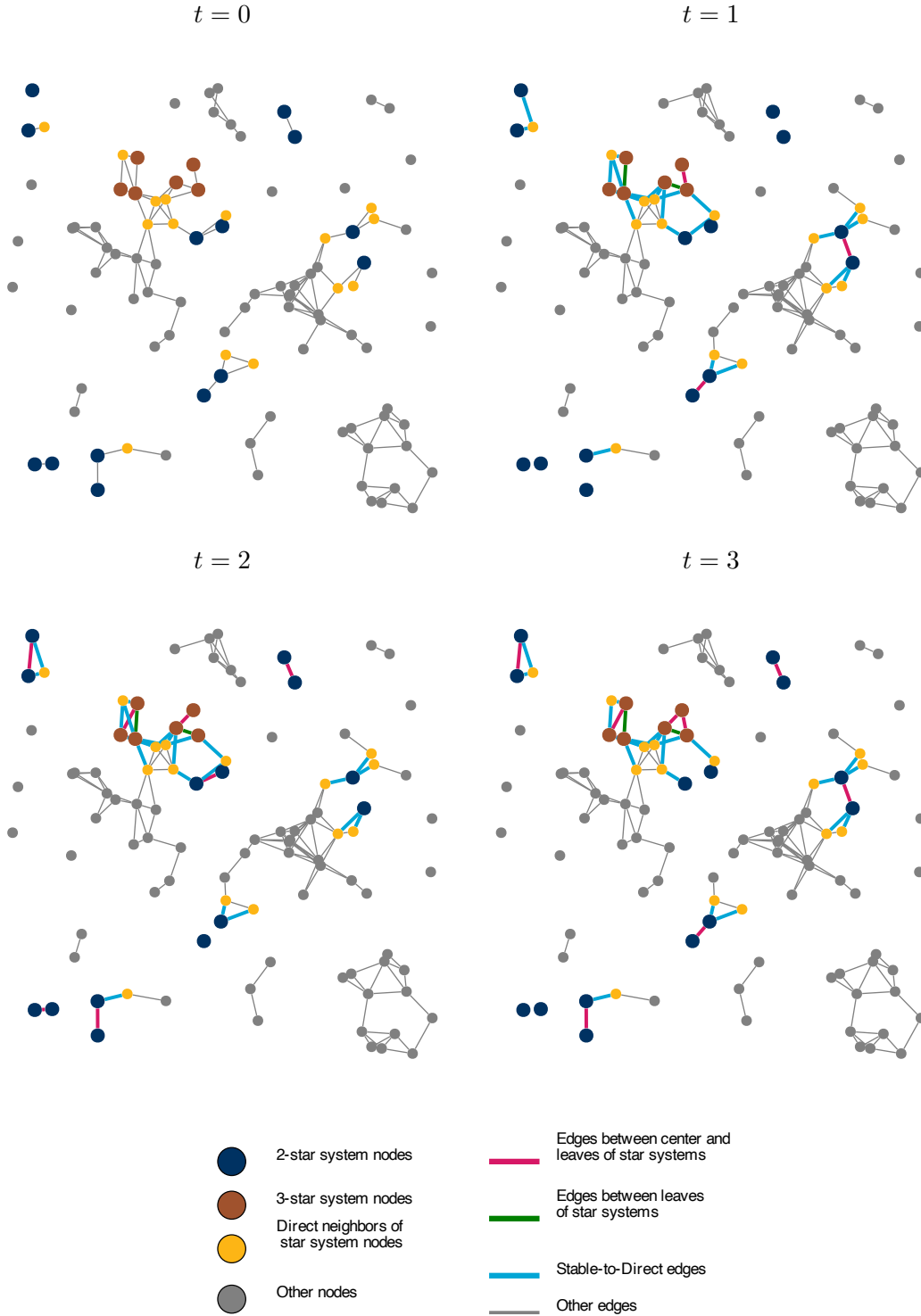
Table 2: Sampling properties of the conditional star systems estimator

Asymptotic Degree	2		3		4	
$N = 5000$	α	β	α	β	α	β
Panel A: Conditional 2-star systems estimator						
Mean Bias	0.0036	0.0326	0.0290	0.0443	0.0562	0.0400
Median Bias	-0.0020	0.0040	0.0090	0.0156	0.0266	0.0168
Mean Abs. Err.	0.1835	0.1986	0.2377	0.2003	0.3169	0.2096
Std. Dev.	0.2313	0.2554	0.2954	0.2525	0.4028	0.2688
Mean Std. Err.	0.2315	0.2448	0.2971	0.2442	0.3926	0.2608
Coverage (95% CI)	0.9540	0.9550	0.9610	0.9490	0.9640	0.9610
Panel B: Conditional 2 & 3-star systems estimator						
Mean Bias	0.0035	0.0289	0.0295	0.0298	0.0399	0.0350
Median Bias	0.0083	0.0073	0.0162	0.0145	0.0102	0.0172
Mean Abs. Err.	0.1633	0.1717	0.2011	0.1672	0.2738	0.1832
Std. Dev.	0.2015	0.2169	0.2510	0.2112	0.3461	0.2326
Mean Std. Err.	0.1993	0.2079	0.2513	0.2084	0.3295	0.2271
Coverage (95% CI)	0.9550	0.9530	0.9590	0.9560	0.9520	0.9590
Avg. # of Stable 2-star systems	290.0		203.3		136.8	
Avg. # of Stable 3-star systems	59.5		47.1		32.2	

Notes: The table reports the sampling properties of $\hat{\theta}$ (using 2-star systems in Panel A, and then 2-star and 3-star systems together in Panel B) across the $B = 1,000$ Monte Carlo simulations for each design ($N = 5,000$). See the main text for other design details.

²²In light of the relatively small number of available 3-stars under this DGP, we have not searched for additional identifying star systems of larger dimensions.

Figure 6: Illustrative network sequence from Monte Carlo Design



Notes: A typical sequence of networks for the case where $N = 100$ and average degree in the initial network is 2.

5 Empirical Application: Co-authorship Networks

Ductor et al. (2014b) studied how researchers’ co-authorship network helps predict their future research output. To illustrate our methodology in a concrete setting, we repurpose their data – assembled from EconLit, a bibliography of journal articles compiled by the American Economic Association – and model the choice of co-authors directly.²³ The Ductor et al. (2014b) datasets spans the period 1970-1999 and is restricted to papers with at most three authors. We focus on the most recent four-year window available: $t = 1996, 1997, 1998, 1999$. For each year t and each pair of authors i and j , we set $D_{ijt} = 1$ if authors i and j co-authored at least one paper in year t , and set $D_{ijt} = 0$ otherwise.

Descriptive statistics for the co-authorship network sequence are detailed in Table 3. Our sample size is $N = 56,156$ authors. The networks are fairly sparse with averages degrees ranging between 0.36 and 0.43 and a share of isolates consistently above 70%. This structure coincides with a *low* average degree random geometric graph; each of our four networks is fragmented into many small connected components.

In the data, many of the observed triangles are due to triple-authored papers (versus three separate bilateral collaborations). Whether our statistical model appropriately captures the micro-foundations of collaborator choices is doubtful. Our goal is simply to provide a worked example on a real world dataset as a “proof-of-concept”.

Table 3: Basic properties of the co-authorship network over 1996-1999

Period	# of authors	Average degree	Density	Share Isolated	T	GC
$t = 0$	56156	0.36	6.3e-6	0.78	0.48	0.0010
$t = 1$	56156	0.37	6.6e-6	0.77	0.48	0.0024
$t = 2$	56156	0.40	7.1e-6	0.75	0.48	0.0019
$t = 3$	56156	0.43	7.7e-6	0.73	0.50	0.0013

Notes: The second column reports average degree or the average number of collaborations per author, the third column documents network density or the share of realized collaborations, the fourth column indicates the share of authors with no collaborations, T is the global clustering coefficient or transitivity index, and GC denotes the fraction of agents that are part of the largest component.

We report four estimates of α and β . First, we fit a standard logistic regression of D_{ijt} on D_{ijt-1} and R_{ijt-1} , without controlling for fixed effects. We then report three versions of our conditional star-system likelihood estimator, based respectively on: (i) identifying 2-star systems only, (ii) identifying 2 and 3-star systems, and (iii) identifying 2, 3, and 4-star systems. The results are collected in Table 4.

²³The data that support the findings of this study are openly available on the *Harvard Dataverse*, DOI: <https://doi.org/10.7910/DVN/27138>.

A first takeaway from Table 4 is that our co-authorship network features a sizable number of identifying star systems, including 3798 identifying 2-star systems. This represents a fairly large effective sample size for our estimators and supports reliable inference in this empirical setting. Their number, however, declines rapidly with system size: from 3798 identifying 2-star systems to 560 identifying 3-star systems, and only 61 identifying 4-star systems. Consequently, incorporating stable 4-systems in the last column of Table 4 produces very modest changes in parameter estimates and standard errors relative to solely relying on identifying 2 and 3-star systems (Column 3).

The logistic regression results suggests that sequences of the form $D_{ikt-1} = 1, D_{jkt-1} = 1, D_{ijt} = 1$ (i.e., instances of links forming to close an open two-star from the previous period) are relatively infrequent in the data. This is indicated by the negative estimate of β . In contrast, the positive estimate of α indicates an overall high-level of bi-lateral link persistence. Some portion of these patterns simply reflects the weaker persistence of three-way collaborations across periods compared to two-way ones.

Columns 2 to 4 report CSS-MLE estimates. In general, these estimates suggest that the high transitivity index observed (~ 0.5 in Table 3) is mainly driven by homophily rather than an intrinsic preference for transitive triplets. Furthermore, the coefficient in the standard logit regression is considerably larger (less negative), indicating that the logit model overstates the role of transitivity by failing to account for homophily-driven clustering.

Table 4: Conditional star system likelihood estimates for the period 1996-1999

	Logit	2-star MLE	2&3-star MLE	2&3&4-star MLE
$\hat{\alpha}$	2.798 (0.096)	0.251 (0.111)	0.357 (0.092)	0.362 (0.089)
$\hat{\beta}$	-0.119 (0.035)	-0.656 (0.312)	-0.564 (0.204)	-0.511 (0.198)
# of authors	56156	56156	56156	56156
# of identifying 2-star systems		3798	3798	3798
# of identifying 3-star systems			560	560
# of identifying 4-star systems				61
# of relevant star systems		365	532	555

Notes: The last row of the table indicates the number of identifying star systems that have a non zero contribution to the conditional star system likelihood. Asymptotic standard errors are reported in parentheses.

6 Conclusion

The analysis of this paper suggests that simple models of dynamic network formation with rich heterogeneity structures can be identified from a single network sequence. Furthermore, consistent estimation, as well as large single network inference, is feasible in some settings. "Watch out, you might get what you're after." Our reliance on logit errors, as in related single agent panel settings, is consequential. At the same time, we expect aspects of our analysis to extend beyond this assumption.

Our work suggests several open questions. First, questions of (asymptotic) efficiency and information accumulation are delicate in our setting. The maximal rate-of-convergence for our conditional star-system maximum likelihood estimator is $\sqrt{N}$. Recall that we observe $\binom{N}{2} = O(N^2)$ sequences of link decisions; so a $\sqrt{N}$ rate-of-convergence is analogous to a $N^{\frac{1}{4}}$ rate in the traditional single-agent panel setting. Our examples suggest that achieving this rate depends upon features of the heterogeneity distribution and initial network configuration. In practice researchers can count the number of identifying star-systems in their dataset. As noted early, with θ consisting of just two parameters, and the CSS likelihood being globally concave, we expect our asymptotic theory to be a good guide to finite-network properties even in settings with only a modest number of identifying star-systems. Whether such star-systems are present in real-world networks, however, is an open question. Our empirical illustration is hopeful in this regard.

Our CSS-MLE estimator utilizes only a subset of the information contained in the "full" conditional likelihood. What is the nature of this information loss? Also of interest is understanding the differential information content of star-systems of different sizes.

It would also be of interest to study the identifiability of other models of link formation. Our focus on transitivity is inspired by empirical research on social and economic networks, but other types of preferences over (local) network structure are of interest as well (see Pelican and Graham (2026)). Other questions involve the inclusion of time-varying regressors and the identifiability of average effects (cf., Dano (2023), Dano et al. (2025), Gao and Niu (2026)).

Data Availability Statement

The data used in this paper were extracted from a larger dataset assembled by Ductor et al. (2014b) and available in the associated Harvard Dataverse repository (Ductor et al., 2014a). We certify that the author(s) of the manuscript have legitimate access to, and permission to use, the data used in this manuscript.

References

- Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society B*, 32(2):283 – 301.
- Bloch, F. and Jackson, M. O. (2006). Definitions of equilibrium in network formation games. *International Journal of Game Theory*, 34(3):305 – 318.
- Card, D. and Sullivan, D. (1988). Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica*, 56(3):497 – 530.
- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225 – 238.
- Chamberlain, G. (1985). *Longitudinal Analysis of Labor Market Data*, chapter Heterogeneity, omitted variable bias, and duration dependence, pages 3 – 38. Cambridge University Press, Cambridge.
- Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press, Cambridge, MA.
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society B*, 20(2):215 – 241.
- Dano, K. (2023). Transition probabilities and moment restrictions in dynamic fixed effects logit models. *arXiv preprint arXiv:2303.00083*.
- Dano, K., Honoré, B. E., and Weidner, M. (2025). Binary choice logit models with general fixed effects for panel and network data. *arXiv preprint arXiv:2508.11556*.
- Dekker, D., Krackhardt, D., and Snijders, T. A. B. (2019). Transitivity correlation. *Network Science*, 7(3):353 – 375.
- Ductor, L., Fafchamps, M., Goyal, S., and van der Leij, M. (2014a). Replication data for: Social Networks and Research Output. Harvard Dataverse. <https://doi.org/10.7910/DVN/27138>.
- Ductor, L., Fafchamps, M., Goyal, S., and van der Leij, M. J. (2014b). Social networks and research output. *The Review of Economics and Statistics*, 96(5):936–948.
- Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326 – 350.
- Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.

- Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87 – 94.
- Gao, W. Y. and Niu, Y. (2026). Identification in dynamic dyadic network formation models with fixed effects. *arXiv preprint arXiv:2604.07488*.
- Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business and Economic Statistics*, 31(3):253 – 264.
- Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033 – 1063.
- Graham, B. S. (2020). *Handbook of Econometrics*, volume 7, chapter Network data, pages 111 – 218. North-Holland, Amsterdam, 1st edition.
- Graham, B. S. and Pelican, A. (2020). *The Econometric Analysis of Network Data*, chapter Testing for externalities in network formation using simulation, pages 65 – 82. Academic Press, London.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, New York.
- Harary, F. and Kimmel, H. J. (1979). Matrix measures for transitivity and balance. *Journal of Mathematical Sociology*, 6(2):199 – 210.
- Heckman, J. J. (1978). Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales de l'insée*, 30-31:227 – 270.
- Heckman, J. J. (1981a). *Structural Analysis of Discrete Data and Econometric Applications*, chapter Statistical models for discrete panel data,, pages 114 – 178. The MIT Press.
- Heckman, J. J. (1981b). *Structural Analysis of Discrete Data and Econometric Applications*, chapter The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process, pages 179 – 195. The MIT Press.
- Heckman, J. J. (1981c). *Studies in Labor Markets*, chapter Heterogeneity and state dependence, pages 91 – 139. Chicago University Press.
- Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839 – 874.
- Honoré, B. E. and Tamer, E. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):611 – 629.
- Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press, Princeton.

- Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. (2012). Social capital and social quilts: network patterns of favor exchange. *American Economic Review*, 102(5):1857–1897.
- Kostin, A. (2010). Probability distribution of distance between pairs of nearest stations in wireless network. *Electronics Letters*, 46(18):1299–1300.
- Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(2):204 – 213.
- McDonald, J. W., Smith, P. W. F., and Forster, J. J. (2007). Markov chain monte carlo exact inference for social networks. *Social Networks*, 29(1):127 – 136.
- McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *The Annals of Probability*, 2(4):620–628.
- McPherson, M., Smith-Lovin, L., and Brashears, M. E. (2006). Social isolation in america: changes in core discussion networks over two decades. *American Sociological Review*, 71(3):353 – 375.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27(1):415 – 444.
- Mitzenmacher, M. and Upfal, E. (2005). *Probability and Computing*. Cambridge University Press, Cambridge.
- Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59(4):1161–1167.
- Newey, W. K. and McFadden, D. (1994). *Handbook of Econometrics*, volume 4, chapter Large sample estimation and hypothesis testing, pages 2111 – 2245. North-Holland, Amsterdam.
- Pelican, A. and Graham, B. S. (2026). An optimal test for strategic interaction in network formation games. Technical report, National Bureau of Economic Research.
- Penrose, M. (2003). *Random Geometric Graphs*. Oxford University Press, Oxford.
- Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, 37:131 – 153.
- van Duijn, M. A. J., Snijders, T. A. B., and Zijlstra, B. J. H. (2004). p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2):234 – 254.

Supplemental Web Appendices

These appendices include proofs of the theorems stated in the main text, as well as statements and proofs of supplemental lemmata. All notation is as established in the main text, unless stated otherwise. Equation numbering continues in sequence with that established in the main text.

A Preliminary results

In this Appendix we prove some of the Lemmas stated in the main text. We also state (and prove as needed) several supplemental results which underlie the proofs of the main Theorems, which appear in Appendix B below.

Throughout a C denotes a generic constant, the magnitude of which may vary throughout. The abbreviations TI, CSI, MI, HI and JI respectively denote Triangle Inequality, Cauchy-Schwarz Inequality, Hoeffding, Markov Inequality, and Jensen's Inequality. The Law of Iterated Expectations is abbreviated by LIE.

A.1 Preliminary results

To prove Theorem 2, stated in the main text, we utilize a textbook result on finite Markov chains (e.g., Theorem 7.10 on p. 172 of Mitzenmacher and Upfal (2005)).

Theorem 5. (STATIONARY DISTRIBUTION) *Consider a finite, irreducible, and ergodic Markov chain with L states and transition matrix $\mathbf{P} = [p_{lm}]_{1 \leq l, m \leq L}$. If (i) there exists a vector of non-negative numbers $\underline{\pi} = (\pi_1, \dots, \pi_L)'$ such that (i) $\sum_{l=1}^L \pi_l = 1$ and (ii) $\pi_l p_{lm} = \pi_m p_{ml}$ for any pair $l, m = 1, \dots, L$, then $\underline{\pi}$ is the stationary distribution of the Markov chain with transition matrix $\mathbf{P}$.*

To demonstrate consistency of the conditional star-system MLE, Theorem 3 in the main text, we invoke a variant of the standard M-Estimation consistency theorem (e.g. Theorem 2.1 in Newey and McFadden (1994) or Theorem 4.1 in Amemiya (1985)). We present and prove the precise result we use here.

Lemma 5. (CONSISTENCY UNDER UNIFORM APPROXIMATION AND IDENTIFICATION) *Let $\Theta \subset \mathbb{R}^{\dim(\theta)}$ be compact. For each N , let $\tilde{Q}_N : \Theta \rightarrow \mathbb{R}$ and $\hat{Q}_N : \Theta \rightarrow \mathbb{R}$ be two stochastic*

functions. Define the estimator

$$\hat{\theta}_N \in \arg \max_{\theta \in \Theta} \hat{Q}_N(\theta).$$

Assume:

A Uniform (in θ) approximation of the target:

$$\sup_{\theta \in \Theta} |\hat{Q}_N(\theta) - \tilde{Q}_N(\theta)| \xrightarrow{p} 0.$$

(For example, via a uniform law of large numbers or stochastic equicontinuity.)

B Well-separated maximizer (i.e., identification assumption): There exists $\theta_0 \in \Theta$ such that for every $\varepsilon > 0$,

$$\Pr \left(\liminf_{N \rightarrow \infty} \left\{ \tilde{Q}_N(\theta_0) - \sup_{\|\theta - \theta_0\| \geq \varepsilon} \tilde{Q}_N(\theta) \right\} > 0 \right) = 1.$$

Then

$$\hat{\theta}_N \xrightarrow{p} \theta_0.$$

Proof. Fix $\varepsilon > 0$. Under the event $\mathbf{1} \left(\liminf_{N \rightarrow \infty} \left\{ \tilde{Q}_N(\theta_0) - \sup_{\|\theta - \theta_0\| \geq \varepsilon} \tilde{Q}_N(\theta) \right\} > 0 \right)$, for large N there exists $\eta > 0$ such that

$$\tilde{Q}_N(\theta_0) \geq \sup_{\|\theta - \theta_0\| \geq \varepsilon} \tilde{Q}_N(\theta) + 3\eta.$$

If $\sup_{\Theta} |\hat{Q}_N - \tilde{Q}_N| \leq \eta$, then

$$\hat{Q}_N(\theta_0) \geq \sup_{\|\theta - \theta_0\| \geq \varepsilon} \hat{Q}_N(\theta) + \eta,$$

so any maximizer of $\hat{Q}_N$ lies within ε of θ_0 . Since the events $\mathbf{1} \left(\liminf_{N \rightarrow \infty} \left\{ \tilde{Q}_N(\theta_0) - \sup_{\|\theta - \theta_0\| \geq \varepsilon} \tilde{Q}_N(\theta) \right\} > 0 \right)$ and $\mathbf{1} \left(\sup_{\Theta} |\hat{Q}_N - \tilde{Q}_N| \leq \eta \right)$ occur with probability approaching one by A and B, it follows that $\hat{\theta}_N \xrightarrow{p} \theta_0$. $\square$

A.2 Proofs of preliminary results appearing in the main text

Proof of Lemma 3 (PERMUTATIONS)

Proof. Part (A): Generically we have, for any $ij \in \mathcal{S}_p$ (with $\mathcal{S}_p$ an identifying p -star system with i at its center) ,

$$\begin{aligned}
R_{ij2} - R_{ij1} &= \sum_{l=1}^N (D_{il2}D_{jl2} - D_{il1}D_{jl1}) \\
&= \sum_{k \in v(\mathcal{S}_p)} (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) + \sum_{l \notin v(\mathcal{S}_p)} (D_{il2}D_{jl2} - D_{il1}D_{jl1}) \\
&= \sum_{k \in v(\mathcal{S}_p)} (D_{ik2}D_{jk2} - D_{ik1}D_{jk1}) + 0 \\
&= \sum_{k \in v(\mathcal{S}_p)} (D_{ik2} - D_{ik1}) D_{jk1}, \tag{23}
\end{aligned}$$

or, equivalently,

$$R_{ij2} = R_{ij1} + \sum_{k \in v(\mathcal{S}_p)} (D_{ik2} - D_{ik1}) D_{jk1}.$$

The third and fourth equalities in (23) follow, respectively from parts (i) and (iii) of Definition 4. Next observe that, since i is the center of $\mathcal{S}_p$, we have for any $\mathbf{B}^3 \in \mathbb{B}_K(\mathbf{D}^3)$, either $(B_{ik1}, B_{ik2}) = (D_{ik1}, D_{ik2})$ for all $k \in v(\mathcal{S}_p)$ or $(B_{ik1}, B_{ik2}) = (D_{ik2}, D_{ik1})$ for all $k \in v(\mathcal{S}_p)$. The definition of $\mathbb{B}_K(\mathbf{D}^3)$ then implies:

$$\begin{aligned}
S_{ij1} &= \sum_{k \in v(\mathcal{S}_p)} B_{ik1}D_{jk1} + \sum_{l \notin v(\mathcal{S}_p)} D_{il1}D_{jl1} \\
&= \sum_{k \in v(\mathcal{S}_p)} (B_{ik1}D_{jk1} - D_{ik1}D_{jk1}) + \sum_{l \notin v(\mathcal{S}_p)} D_{il1}D_{jl1} + \sum_{k \in v(\mathcal{S}_p)} D_{ik1}D_{jk1} \\
&= \sum_{k \in v(\mathcal{S}_p)} (B_{ik1}D_{jk1} - D_{ik1}D_{jk1}) + \sum_l D_{il1}D_{jl1} \\
&= \sum_{k \in v(\mathcal{S}_p)} (B_{ik1} - D_{ik1}) D_{jk1} + R_{ij1} \\
&= \sum_{k \in v(\mathcal{S}_p)} (D_{ik2} - D_{ik1}) D_{jk1} - \sum_{k \in v(\mathcal{S}_p)} (D_{ik2} - B_{ik1}) D_{jk1} + R_{ij1} \\
&= R_{ij2} - R_{ij1} - \sum_{k \in v(\mathcal{S}_p)} (D_{ik2} - B_{ik1}) D_{jk1} + R_{ij1} \\
&= R_{ij2} - \sum_{k \in v(\mathcal{S}_p)} (D_{ik2} - B_{ik1}) D_{jk1}. \tag{24}
\end{aligned}$$

Hence, if $(B_{ik1}, B_{ik2}) = (D_{ik1}, D_{ik2})$ for all $k \in v(\mathcal{S}_p)$, we have, using (23) and (24), that $S_{ij1} = R_{ij1}$ (and, by symmetry, that $S_{ij2} = R_{ij2}$). If $(B_{ik1}, B_{ik2}) = (D_{ik2}, D_{ik1})$ for all $k \in v(\mathcal{S}_p)$, (24) instead gives $S_{ij1} = R_{ij2}$ (and, by symmetry, that $S_{ij2} = R_{ij1}$).

Part (B): Generically we have, for any $(j, k) \in v(\mathcal{S}_p) \times v(\mathcal{S}_p)$ such that $jk \in \mathcal{D}_p^{fd}$

$$\begin{aligned}
R_{jk2} - R_{jk1} &= \sum_{l=1}^N (D_{jl2}D_{kl2} - D_{jl1}D_{kl1}) \\
&= \sum_{l \in v(\mathcal{S}_p)} (D_{jl2}D_{kl2} - D_{jl1}D_{kl1}) + \sum_{m \notin v(\mathcal{S}_p)} (D_{jm2}D_{km2} - D_{jm1}D_{km1}) \\
&= \sum_{l \in v(\mathcal{S}_p)} (D_{jl2}D_{kl2} - D_{jl1}D_{kl1}) + 0 \\
&= D_{ij2}D_{ik2} - D_{ij1}D_{ik1} + \sum_{l \in v(\mathcal{S}_p), l \neq i} (D_{jl2}D_{kl2} - D_{jl1}D_{kl1}) \\
&= D_{ij2}D_{ik2} - D_{ij1}D_{ik1} + 0.
\end{aligned} \tag{25}$$

The third and fifth equalities above follow from, respectively, parts (i) and (iii) of Definition 4. Next observe that, invoking parts (i) and (iii) of Definition 4, and also using (25),

$$\begin{aligned}
S_{jk1} &= B_{ij1}B_{ik1} + \sum_{l \neq i} D_{jl1}D_{kl1} \\
&= B_{ij1}B_{ik1} + \sum_{l \neq i} D_{jl2}D_{kl2} \\
&= R_{jk2} - (D_{ij2}D_{ik2} - B_{ij1}B_{ik1}).
\end{aligned}$$

Hence if $(B_{ij1}, B_{ij2}) = (D_{ij1}, D_{ij2})$ and $(B_{ik1}, B_{ik2}) = (D_{ik1}, D_{ik2})$, then $S_{jk1} = R_{jk1}$ (and, by symmetry, $S_{jk2} = R_{jk2}$; the first equality above). If, instead, $(B_{ij1}, B_{ij2}) = (D_{ij2}, D_{ij1})$ and $(B_{ik1}, B_{ik2}) = (D_{ik2}, D_{ik1})$, then $S_{jk1} = R_{jk2}$ (and, by symmetry, $S_{jk2} = R_{jk1}$; the last equality above).

Part (C): Recall that $j \in v(\mathcal{S}_p)$ and $l \notin v(\mathcal{S}_p)$. There are two cases to consider. In case (i), while agent l is not part of $v(\mathcal{S}_p)$, it's in its period 1 neighborhood (i.e., $l \in n_1(\mathcal{S}_p)$). In case (ii) it is not in this neighborhood (i.e., $l \notin n_1(\mathcal{S}_p)$). In the first case, invoking parts (i) to (iii) of Definition 4 (such that $B_{jm1} = D_{jm1}$ for $m \neq i$ and $B_{lm1} = D_{lm1}$ for all $m = 1, \dots, N$), we have:

$$\begin{aligned}
S_{jl1} &= B_{ij1}D_{il1} + \sum_{m \neq i} D_{jm1}D_{lm1} \\
&= B_{ij1}D_{il1} - D_{ij1}D_{il1} + R_{jl1} \\
&= (B_{ij1} - D_{ij1})D_{il1} + R_{jl1}.
\end{aligned}$$

Hence if $B_{ij1} = D_{ij1}$ we have that $S_{jl1} = R_{jl1}$ (and, by symmetry, $S_{jl2} = R_{jl2}$). If, instead,

$B_{ij1} = D_{ij2}$, we have (again invoking the parts (i) and (ii) of Definition 4)

$$\begin{aligned}
S_{jl1} &= B_{ij1}D_{il1} + \sum_{m \neq i} D_{jm1}D_{lm1} \\
&= B_{ij1}D_{il1} + \sum_{m \neq i} D_{jm2}D_{lm2} \\
&= B_{ij1}D_{il1} - D_{ij2}D_{il2} + R_{jl2} \\
&= (B_{ij1} - D_{ij2})D_{il2} + R_{jl2},
\end{aligned}$$

such that $S_{jl1} = R_{jl2}$ (and, by symmetry, $S_{jl2} = R_{jl1}$).

Next consider case (ii) where $l \notin n_1(\mathcal{S}_p)$. There are two sub-cases. If $l \in v(\mathcal{S}_l)$ for some other identifying star system with center i' then

$$\begin{aligned}
S_{jl1} &= B_{ij1}D_{il1} + D_{i'j1}B_{i'l1} + \sum_{m \neq i, i'} D_{jm1}D_{lm1} \\
&= B_{ij1}D_{il1} + D_{i'j1}B_{i'l1} + \sum_{m \neq i, i'} D_{jm2}D_{lm2},
\end{aligned}$$

by part (i) of Definition 4. Next, since $D_{il1} = D_{il2} = D_{i'j1} = D_{i'j2} = 0$ by Lemma 2, we can use the argument from case (i) above to conclude that $S_{il1} = S_{il2} = R_{il1} = R_{il2}$ and the conclusion trivially holds. In the second sub-case l is neither in the neighborhood of $\mathcal{S}_p$ nor does it belong to another identifying star system. The link structure of agent l is invariant across the elements of $\mathbf{B}^3 \in \mathbb{B}_K(\mathbf{D}^3)$ and we can show the result by an adaptation of the argument used for case (i). $\square$

Proof of Lemma 4 (CONDITIONAL INDEPENDENCE)

Proof. For any network sequence $\tilde{\mathbf{d}}$, denote $Z_{pi}(\tilde{\mathbf{d}})$ the indicator of whether $\mathbf{i}$ is stable in $\tilde{\mathbf{d}}$ and $\mathbf{Q}_{pi}(\tilde{\mathbf{d}})$ the tuple $\mathbf{Q}_{pi}$ defined on $\tilde{\mathbf{d}}$. Let $\mathbf{Q}_{-\mathbf{Z}}(\tilde{\mathbf{d}})$ collect all remaining dyads Q_{ij} that do not appear in any $\mathbf{Q}_{pi}(\tilde{\mathbf{d}})$. Write the joint probability mass function of the tuples $(Z_{pi}\mathbf{Q}_{pi})_{p \leq K, \mathbf{i}}, \mathbf{Q}_{-\mathbf{Z}}$:

$$\Pr((Z_{pi}\mathbf{Q}_{pi})_{p \leq K, \mathbf{i}}, \mathbf{Q}_{-\mathbf{Z}} | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}) = \sum_{\tilde{\mathbf{d}}: \forall p, \mathbf{i} : \mathbf{Q}_{pi}(\tilde{\mathbf{d}}) = \mathbf{Q}_{pi}; \mathbf{Q}_{-\mathbf{Z}}(\tilde{\mathbf{d}}) = \mathbf{Q}_{-\mathbf{Z}}} \Pr(\tilde{\mathbf{d}} | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}).$$

Note that for a fixed $\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}$, the map $\tilde{\mathbf{d}} \rightarrow [(Z_{pi}(\tilde{\mathbf{d}})\mathbf{Q}_{pi}(\tilde{\mathbf{d}}))_{p, \mathbf{i}}, \mathbf{Q}_{-\mathbf{Z}}(\tilde{\mathbf{d}})]$ is a bijection,

therefore:

$$\begin{aligned} \Pr((Z_{pi}\mathbf{Q}_{pi})_{p \leq K, i}, \mathbf{Q}_{-Z} | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}) &= \Pr(\mathbf{d} | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}) \\ &= \prod_{p=2}^K \prod_i \left(\frac{\exp(T'_{pi}\theta)}{1 + \exp(T'_{pi}\theta)} \right)^{Z_{pi}}, \end{aligned}$$

by Theorem 1. Let $c(\mathbf{Z}, \mathbf{D}_0, \mathbf{D}_3)$ be the number of all potential values of $\mathbf{Q}_{-Z}$, summing over all of the potential values of $\mathbf{Q}_{-Z}$, we get:

$$\Pr((Z_{pi}\mathbf{Q}_{pi})_{p \leq K, i} | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}) = c(\mathbf{Z}, \mathbf{D}_0, \mathbf{D}_3) \cdot \prod_{p=2}^K \prod_i \left(\frac{\exp(T'_{pi}\theta)}{1 + \exp(T'_{pi}\theta)} \right)^{Z_{pi}},$$

implying that $(Z_{pi}\mathbf{Q}_{pi})_{p \leq K, i}$ are independent conditionally on $\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}$, with a marginal:

$$\Pr(Z_{pi}\mathbf{Q}_{pi} | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}) \propto \left(\frac{\exp(T'_{pi}\theta)}{1 + \exp(T'_{pi}\theta)} \right)^{Z_{pi}}.$$

□

B Main results

B.1 Proof of Proposition 1

Part (i) follows from inspection of (9) and (11). To show (ii) let $\mathbf{b}^3 \in \mathbb{B}(\mathbf{d}^3)$: because $\mathbf{b}_0 = \mathbf{d}_0$, $\mathbf{b}_3 = \mathbf{d}_3$, and $\mathbf{b}_1 + \mathbf{b}_2 = \mathbf{d}_1 + \mathbf{d}_2$, then $\sum_{t=1}^3 b_{ijt} = \sum_{t=1}^3 d_{ijt}$. Moreover, fix a pair ij , $\delta \in \{0, 1\}$ and $\rho \in \{0, 1, \dots, N-2\}$. Since $\mathbf{b}_0 = \mathbf{d}_0$, then $1 \in m_{ij}(\delta, \rho; \mathbf{d}^3)$ if, and only if, $1 \in m_{ij}(\delta, \rho; \mathbf{b}^3)$. Given that $(s_{ij1}, s_{ij2}) = (r_{ij1}, r_{ij2})$ if $(b_{ij1}, b_{ij2}) = (d_{ij1}, d_{ij2})$ and $(s_{ij1}, s_{ij2}) = (r_{ij2}, r_{ij1})$ if $(b_{ij1}, b_{ij2}) = (d_{ij2}, d_{ij1})$, then $|m_{ij}(\delta, \rho; \mathbf{b}^3) \cap \{2, 3\}| = |m_{ij}(\delta, \rho; \mathbf{d}^3) \cap \{2, 3\}|$, therefore $|m_{ij}(\delta, \rho; \mathbf{b}^3)| = |m_{ij}(\delta, \rho; \mathbf{d}^3)|$.

Conversely, let $\mathbf{b}^3 \in \mathbb{B}^*$ and note that $d_{ij0} + d_{ij1} + d_{ij2} = \sum_{\rho=0}^{N-2} m_{ij}(1, \rho; \mathbf{d}^3) = \sum_{\rho=0}^{N-2} m_{ij}(1, \rho; \mathbf{b}^3) = b_{ij0} + b_{ij1} + b_{ij2}$. Since: $\mathbf{b}_0 = \mathbf{d}_0$ by assumption, we get: $d_{ij1} + d_{ij2} = b_{ij1} + b_{ij2}$. The fact that $\sum_{t=1}^3 b_{ijt} = \sum_{t=1}^3 d_{ijt}$ for all i, j allows us to conclude that $\mathbf{d}_3 = \mathbf{b}_3$.

B.2 Proof of Theorem 1 (Conditional Star-System Likelihood)

In order to simplify (17) it is convenient to analyze its inverse, which consists of the sum of $2^{\sum_{p=2}^K \mathbf{m}_{p,N}}$ ratios of the form

$$\frac{\Pr(\mathbf{D}^3 = \mathbf{b}^3; \theta, \mathbf{A}, \pi)}{\Pr(\mathbf{D}^3 = \mathbf{d}^3; \theta, \mathbf{A}, \pi)}, \quad (26)$$

with $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$. Note that the cardinality of the set $\mathbb{B}_K(\mathbf{d}^3)$ is $2^{\sum_{p=2}^K \mathbf{m}_{p,N}}$.

From the dyad partition given in equation (15) above, we can decompose the denominator of (26) as follows:

$$\Pr(\mathbf{d}^3; \theta, \mathbf{A}, \pi) = \pi(\mathbf{d}_0; \mathbf{A}) \quad (27)$$

$$\begin{aligned} & \times \prod_{p=2}^K \prod_{ij \in \mathcal{D}_p} \prod_{t=1}^3 F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})^{d_{ij}} \\ & \times [1 - F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})]^{1-d_{ij}} \end{aligned} \quad (28)$$

$$\begin{aligned} & \times \prod_{p=2}^K \prod_{ij \in \mathcal{D}_p^{fd}} \prod_{t=1}^3 F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})^{d_{ij}} \\ & \times [1 - F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})]^{1-d_{ij}} \end{aligned} \quad (29)$$

$$\begin{aligned} & \times \prod_{p=2}^K \prod_{ij \in \mathcal{D}_{p,o}^{pd}} \prod_{t=1}^3 F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})^{d_{ij}} \\ & \times [1 - F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})]^{1-d_{ij}} \end{aligned} \quad (30)$$

$$\begin{aligned} & \times \prod_{p=2}^K \prod_{l=2}^K \prod_{ij \in \mathcal{D}_{p,l}^{pd}} \prod_{t=1}^3 F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})^{d_{ij}} \\ & \times [1 - F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})]^{1-d_{ij}} \end{aligned} \quad (31)$$

$$\begin{aligned} & \times \prod_{ij \in \mathcal{D}_K^{od}} \prod_{t=1}^3 F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})^{d_{ij}} \\ & \times [1 - F(\alpha d_{ijt-1} + \beta r_{ijt-1} + A_{ij})]^{1-d_{ij}} \end{aligned} \quad (32)$$

An analogous decomposition applies to the numerator of (26).

In order to analyze (26) it is convenient use the dyad partition given by (15) and also underlying the likelihood decomposition of equations (27) to (31) immediately above.

Initial conditions. To evaluate (26) we begin with the observation for any $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ we have $\mathbf{b}_0 = \mathbf{d}_0$. Therefore the $\pi(\mathbf{b}_0, \mathbf{a})$ and $\pi(\mathbf{d}_0, \mathbf{a})$ terms in, respectively, the numerator

and denominator of (26) are common factors.

Dyads in $\mathcal{D}_K^{od}$. Next consider the contributions, to the numerator and denominator of (26), of all those dyads where *neither* agent belongs to an identifying star system. These contributions, like those of the initial condition, also coincide. This follows since swapping the edges of an identifying star system leaves the systematic utility associated with (potential) links in $\mathcal{D}_K^{od}$ unchanged. The transitivity term for dyad $kl \in \mathcal{D}_K^{od}$ equals $r_{klt} = \sum_m d_{kmt} d_{lmt}$; none of the links entering this term vary over $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$.

We now turn to evaluating the contributions of dyads where at least one agent belongs to an identifying star system to ratio (26).

Dyads in $\bigcup_{p=2}^K \mathcal{D}_p$. Let $ij \in \mathcal{S}_p$ be a dyad in $\mathcal{S}_p$, an identifying star system of size p . Without loss of generality, suppose that i is the center node of $\mathcal{S}_p$, and consider the periods $t = 1, 2, 3$ contributions of this dyad to (26). There are two cases to consider: that where (i) the edges in $\mathcal{S}_p$ are swapped in $\mathbf{b}^3$ relative to their configuration in $\mathbf{d}^3$, and (ii) where they are not swapped. In the second case, the likelihood contributions of ij to the numerator and denominator of (26) coincide.

If the former case we have either

- (a) $(d_{ij1}, d_{ij2}) = (1, 0)$, such that $(b_{ij1}, b_{ij2}) = (0, 1)$, or
- (b) $(d_{ij1}, d_{ij2}) = (0, 1)$, such that $(b_{ij1}, b_{ij2}) = (1, 0)$.

In subcase (a) ij 's period's $t = 1, 2, 3$ contribution to (26) equals:

$$\begin{aligned}
b^{10}(q_{ij}; \theta) &\stackrel{def}{=} \frac{[1 - F(\alpha b_{ij0} + \beta s_{ij0} + A_{ij})] F(\beta s_{ij1} + A_{ij})}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) [1 - F(\alpha + \beta r_{ij1} + A_{ij})]} \\
&\quad \times \frac{F(\alpha + \beta s_{ij2} + A_{ij})^{b_{ij3}} [1 - F(\alpha + \beta s_{ij2} + A_{ij})]^{1-b_{ij3}}}{F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\
&= \frac{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] F(\beta r_{ij2} + A_{ij})}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) [1 - F(\alpha + \beta r_{ij1} + A_{ij})]} \\
&\quad \times \frac{F(\alpha + \beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\
&= \frac{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{1-d_{ij3}} F(\alpha + \beta r_{ij1} + A_{ij})^{d_{ij3}}}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}} [1 - F(\alpha + \beta r_{ij1} + A_{ij})]^{d_{ij3}}} \\
&= \frac{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] F(\alpha d_{ij3} + \beta r_{ij2} + A_{ij})^{1-d_{ij3}} F(\alpha d_{ij3} + \beta r_{ij1} + A_{ij})^{d_{ij3}}}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) [1 - F(\alpha d_{ij3} + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}} [1 - F(\alpha d_{ij3} + \beta r_{ij1} + A_{ij})]^{d_{ij3}}} \\
&= \frac{1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})} \frac{F(\alpha d_{ij3} + \beta r_{ij1} + A_{ij})}{1 - F(\alpha d_{ij3} + \beta r_{ij1} + A_{ij})} \\
&\quad \times \left(\frac{1 - F(\beta r_{ij1} + A_{ij})}{F(\beta r_{ij1} + A_{ij})} \frac{F(\beta r_{ij2} + A_{ij})}{1 - F(\beta r_{ij2} + A_{ij})} \right)^{1-d_{ij3}}
\end{aligned}$$

where we use $\mathbf{b}_0 = \mathbf{d}_0$, $\mathbf{b}_3 = \mathbf{d}_3$ and Lemma 3 for the second equality and ideas from Honoré and Kyriazidou (2000) to formulate the expression to the right of the final equality.

In the logit case, this ratio simplifies to

$$b^{10}(q_{ij}; \theta) = \exp(\alpha [d_{ij3} - d_{ij0}] + \beta [(r_{ij1} - r_{ij0}) + (1 - d_{ij3})(r_{ij2} - r_{ij1})]),$$

Observe that this expression does not depend on A_{ij} .

In subcase (b), where $(d_{ij1}, d_{ij2}) = (0, 1)$ such that $(b_{ij1}, b_{ij2}) = (1, 0)$, we instead get a

contribution of

$$\begin{aligned}
b^{01}(q_{ij}; \theta) &\stackrel{\text{def}}{=} \frac{F(\alpha b_{ij0} + \beta s_{ij0} + A_{ij}) [1 - F(\alpha + \beta s_{ij1} + A_{ij})]}{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] F(\beta r_{ij1} + A_{ij})} \\
&\quad \times \frac{F(\beta s_{ij2} + A_{ij})^{b_{ij3}} [1 - F(\beta s_{ij2} + A_{ij})]^{1-b_{ij3}}}{F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\
&= \frac{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) [1 - F(\alpha + \beta r_{ij2} + A_{ij})]}{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] F(\beta r_{ij1} + A_{ij})} \\
&\quad \times \frac{F(\beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\
&= \frac{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{d_{ij3}} [1 - F(\beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} F(\beta r_{ij1} + A_{ij})^{1-d_{ij3}}} \\
&= \frac{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})}{1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})} \frac{1 - F(\alpha d_{ij3} + \beta r_{ij1} + A_{ij})}{F(\alpha d_{ij3} + \beta r_{ij1} + A_{ij})} \\
&\quad \times \left(\frac{F(\alpha + \beta r_{ij1} + A_{ij})}{1 - F(\alpha + \beta r_{ij1} + A_{ij})} \frac{1 - F(\alpha + \beta r_{ij2} + A_{ij})}{F(\alpha + \beta r_{ij2} + A_{ij})} \right)^{d_{ij3}}.
\end{aligned}$$

In the logit case we further get

$$b^{01}(q_{ij}; \theta) = \exp(-\alpha [d_{ij3} - d_{ij0}] - \beta [(r_{ij1} - r_{ij0}) + d_{ij3}(r_{ij2} - r_{ij1})]),$$

which again does not vary with A_{ij} . For the analysis which follows, it is convenient – *for the logit case* – to combine $b^{01}(q_{ij}; \theta)$ and $b^{10}(q_{ij}; \theta)$ into a single expression of, for example,

$$\begin{aligned}
b(q_{ij}; \theta) &= \exp \Big(-\alpha (d_{ij2} - d_{ij1})(d_{ij3} - d_{ij0}) \\
&\quad - \beta (d_{ij2} - d_{ij1})(r_{ij1} - r_{ij0}) \\
&\quad + \beta (d_{ij1}(1 - d_{ij2})(1 - d_{ij3}) - (1 - d_{ij1})d_{ij2}d_{ij3})(r_{ij2} - r_{ij1}) \Big).
\end{aligned}$$

This expression can be simplified further by noting that by Definition 5 (ii), since i is the center of the star system, $d_{ij2} = (1 - d_{ij1})$. Hence, after some manipulation we get the form for given in the statement of Theorem 1, equation (18) of the main text.

Dyads in sets $\mathcal{D}_{p,l}^{pd}$. We next turn to dyads composed of agents that belong to two different identifying star-systems. It turns out that for any $(p, l) \in \{2, \dots, K\}^2$, the contributions of dyads $ij \in \mathcal{D}_{p,l}^{pd}$ to the numerator and denominator of (26) coincide. To see this, recall the notation $s_{ijt} = \sum_k b_{ikt} b_{jkt}$ for $\mathbf{b}_t = [b_{ijt}]_{1 \leq i, j \leq N}$, and notice first that by part (A) of Lemma 3, $d_{ij1} = d_{ij2} = 0 = b_{ij1} = b_{ij2}$. Hence, the periods $t = 1, 2, 3$ contributions to (26) for dyad

$ij \in \mathcal{D}_{p,l}^{pd}$ equal

$$\begin{aligned}
& \frac{[1 - F(\alpha b_{ij0} + \beta s_{ij0} + A_{ij})][1 - F(\beta s_{ij1} + A_{ij})] F(\beta s_{ij2} + A_{ij})^{b_{ij3}} [1 - F(\beta s_{ij2} + A_{ij})]^{1-b_{ij3}}}{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})][1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\
&= \frac{[1 - F(\beta s_{ij1} + A_{ij})] F(\beta s_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta s_{ij2} + A_{ij})]^{1-d_{ij3}}}{[1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\
&= \frac{[1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{[1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij1} + A_{ij})]^{1-d_{ij3}}} \\
&= 1
\end{aligned}$$

where the first equality leverages the fact that $\mathbf{b}_0 = \mathbf{d}_0$ and $\mathbf{b}_3 = \mathbf{d}_3$, and the second equality draws on both part (B) of Lemma 3 and Lemma 3 (C).

Dyads in sets $\mathcal{D}_{p,o}^{pd}$. The analysis of dyads with one agent in an identifying star-system and the other not is more complicated. Fix $p \in \{2, \dots, K\}$ and let consider the periods $t = 1, 2, 3$ contributions to (26) for dyad $ij \in \mathcal{D}_{p,o}^{pd}$, where, without loss of generality, we let $i \in \mathcal{V}_p$ and $j \in \mathcal{V}^o$. Let $\mathcal{S}_p$ denote the p -star system to which i belongs. By neighborhood stability of $\mathcal{S}_p$ (Definition 4), $d_{ij1} = d_{ij2}$, and by the definition of $\mathbb{B}_K(\mathbf{d}^3)$, $b_{ij1} = d_{ij1} \wedge b_{ij2} = d_{ij2}$. Therefore there are two cases to consider. In the first, if $(d_{ij1}, d_{ij2}) = (1, 1)$, which yields a likelihood contribution of

$$\begin{aligned}
& \frac{F(\alpha b_{ij0} + \beta s_{ij0} + A_{ij}) F(\alpha + \beta s_{ij1} + A_{ij}) F(\alpha + \beta s_{ij2} + A_{ij})^{b_{ij3}} [1 - F(\alpha + \beta s_{ij2} + A_{ij})]^{1-b_{ij3}}}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) F(\alpha + \beta r_{ij1} + A_{ij}) F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\
&= \frac{F(\alpha + \beta s_{ij1} + A_{ij}) F(\alpha + \beta s_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta s_{ij2} + A_{ij})]^{1-d_{ij3}}}{F(\alpha + \beta r_{ij1} + A_{ij}) F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}}
\end{aligned}$$

where we use $\mathbf{b}_0 = \mathbf{d}_0$ and $\mathbf{b}_3 = \mathbf{d}_3$. If the links of $\mathcal{S}_p$ are not swapped in $\mathbf{b}^3$ relative to their configuration in $\mathbf{d}^3$, then, by Lemma 3, $(s_{ij1}, s_{ij2}) = (r_{ij1}, r_{ij2})$ and the ratio above equals 1. If, instead, the links of $\mathcal{S}_p$ are swapped, then by Lemma 3, the ratio above equals

$$c^{11}(q_{ij}; \theta) \stackrel{def}{=} \frac{F(\alpha + \beta r_{ij2} + A_{ij}) F(\alpha + \beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{F(\alpha + \beta r_{ij1} + A_{ij}) F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}}$$

In the logit case this simplifies further to $c^{11}(q_{ij}; \theta) = \exp(\beta(1 - d_{ij3})(r_{ij2} - r_{ij1}))$. In the second case to consider, if $(d_{ij1}, d_{ij2}) = (0, 0)$ such that the likelihood contribution equals

$$\begin{aligned} & \frac{[1 - F(\alpha b_{ij0} + \beta s_{ij0} + A_{ij})][1 - F(\beta s_{ij1} + A_{ij})] F(\beta s_{ij2} + A_{ij})^{b_{ij3}} [1 - F(\beta s_{ij2} + A_{ij})]^{1-b_{ij3}}}{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})][1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\ &= \frac{[1 - F(\beta s_{ij1} + A_{ij})] F(\beta s_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta s_{ij2} + A_{ij})]^{1-d_{ij3}}}{[1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \end{aligned}$$

Again, if the links of $\mathcal{S}_p$ are not swapped in $\mathbf{b}^3$, then by Lemma 3, $(s_{ij1}, s_{ij2}) = (r_{ij1}, r_{ij2})$ and the ratio equals 1. Otherwise, Lemma 3 implies that it equals

$$c^{00}(q_{ij}; \theta) \stackrel{\text{def}}{=} \frac{[1 - F(\beta r_{ij2} + A_{ij})] F(\beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{[1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}}.$$

In the logit case, this simplified to $c^{00}(q_{ij}; \theta) = \exp(-\beta d_{ij3}(r_{ij2} - r_{ij1}))$. We can – *in the logit case* – combine $c^{11}(q_{ij}; \theta)$ and $c^{00}(q_{ij}; \theta)$ into a single expression of

$$c(q_{ij}; \theta) = \exp(\beta(d_{ij1} - d_{ij3})(r_{ij2} - r_{ij1}))$$

It is convenient for what follows to show, again for the logit case, that $c(q_{ij}; \theta) = b(q_{ij}; \theta)$ for dyad $ij \in \mathcal{D}_{p,o}^{pd}$. Since $d_{ij1} = d_{ij2}$ we have:

$$\begin{aligned} b(q_{ij}; \theta) &= \exp\left(-\alpha(d_{ij2} - d_{ij1})(d_{ij3} - d_{ij0}) - \beta\left([d_{ij3} - d_{ij1}](r_{ij2} - r_{ij0}) - [d_{ij3} - d_{ij2}](r_{ij1} - r_{ij0})\right)\right) \\ &= \exp\left(-\beta\left([d_{ij3} - d_{ij1}](r_{ij2} - r_{ij0}) - [d_{ij3} - d_{ij1}](r_{ij1} - r_{ij0})\right)\right) \\ &= \exp\left(-\beta\left([d_{ij3} - d_{ij1}](r_{ij2} - r_{ij1})\right)\right) \\ &= c(q_{ij}; \theta). \end{aligned}$$

Dyads in sets $\mathcal{D}_p^{fd}$. All that remains to consider are the period $t = 1, 2, 3$ contributions to (26) for those dyads where both agents belong to the same identifying star system but the dyad itself does not (i.e., $ij \in \mathcal{D}_p^{fd}$, with $i, j \in v(\mathcal{S}_p)$ for some p -star identifying system $\mathcal{S}_p$). If the links of $\mathcal{S}_p$ are not swapped in $\mathbf{b}^3$ relative to $\mathbf{d}^3$, then, by Lemma 3, the contributions of ij to the numerator and denominator of (26) coincide. Otherwise, if the links of $\mathcal{S}_p$ are swapped in $\mathbf{b}^3$, there are two cases to consider: (i) $(d_{ij1}, d_{ij2}) = (1, 1) = (b_{ij1}, b_{ij2})$ and (ii)

$(d_{ij1}, d_{ij2}) = (0, 0) = (b_{ij1}, b_{ij2})$. In the first case we have

$$\begin{aligned} & \frac{F(\alpha b_{ij0} + \beta s_{ij0} + A_{ij}) F(\alpha + \beta s_{ij1} + A_{ij}) F(\alpha + \beta s_{ij2} + A_{ij})^{b_{ij3}} [1 - F(\alpha + \beta s_{ij2} + A_{ij})]^{1-b_{ij3}}}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) F(\alpha + \beta r_{ij1} + A_{ij}) F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\ &= \frac{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) F(\alpha + \beta r_{ij2} + A_{ij}) F(\alpha + \beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij}) F(\alpha + \beta r_{ij1} + A_{ij}) F(\alpha + \beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\ &= \frac{[1 - F(\alpha + \beta r_{ij1} + A_{ij})]^{1-d_{ij3}} F(\alpha + \beta r_{ij2} + A_{ij})^{1-d_{ij3}}}{F(\alpha + \beta r_{ij1} + A_{ij})^{1-d_{ij3}} [1 - F(\alpha + \beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \end{aligned}$$

where the first equality follows from $\mathbf{b}_0 = \mathbf{d}_0$, $\mathbf{b}_3 = \mathbf{d}_3$, and Lemma 3. In the logit case, this expression evaluates to $c^{11}(q_{ij}; \theta) = \exp(\beta(1 - d_{ij3})(r_{ij2} - r_{ij1}))$, identical to the case where $ij \in \mathcal{D}_{p,o}^{pd}$. In the second case, we get by similar arguments,

$$\begin{aligned} & \frac{[1 - F(\alpha b_{ij0} + \beta s_{ij0} + A_{ij})] [1 - F(\beta s_{ij1} + A_{ij})] F(\beta s_{ij2} + A_{ij})^{b_{ij3}} [1 - F(\beta s_{ij2} + A_{ij})]^{1-b_{ij3}}}{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] [1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\ &= \frac{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] [1 - F(\beta r_{ij2} + A_{ij})] F(\beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij1} + A_{ij})]^{1-d_{ij3}}}{[1 - F(\alpha d_{ij0} + \beta r_{ij0} + A_{ij})] [1 - F(\beta r_{ij1} + A_{ij})] F(\beta r_{ij2} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{1-d_{ij3}}} \\ &= \frac{F(\beta r_{ij1} + A_{ij})^{d_{ij3}} [1 - F(\beta r_{ij2} + A_{ij})]^{d_{ij3}}}{[1 - F(\beta r_{ij1} + A_{ij})]^{d_{ij3}} F(\beta r_{ij2} + A_{ij})^{d_{ij3}}}. \end{aligned}$$

In the logit case this ratio simplifies to $c^{00}(q_{ij}; \theta) = \exp(-\beta d_{ij3}(r_{ij2} - r_{ij1}))$. As in the *partial dyad* case, we can combine $c^{00}(q_{ij}; \theta)$ and $c^{11}(q_{ij}; \theta)$ into a generic expression of $c(q_{ij}; \theta) = \exp(\beta(d_{ij1} - d_{ij3})(r_{ij2} - r_{ij1}))$. As in our analysis for dyads in $\mathcal{D}_{p,o}^{pd}$, since $d_{ij1} = d_{ij2}$, we also obtain the equivalence $c(q_{ij}; \theta) = b(q_{ij}; \theta)$ for dyads in $\mathcal{D}_p^{fd}$.

The analysis above indicates that the periods $t = 1, 2, 3$ contributions of all dyads either coincide or, in the logit case, equal $b(q_{ij}; \theta)$. The precise contribution of ij depends upon which dyad set it belongs to and whether the edges of various identifying star systems are swapped in $\mathbf{b}^3$ relative to their configuration in $\mathbf{d}^3$.

Let $v(\mathcal{S}_{pi}, \mathbf{b}^3)$ equal 1 if the edges of identifying star system $\mathcal{S}_{pi}$ are swapped and zero otherwise. The analysis above allows us to re-write (26) as.

$$\frac{\Pr(\mathbf{D}^3 = \mathbf{b}^3; \theta, \mathbf{A}, \pi)}{\Pr(\mathbf{D}^3 = \mathbf{d}^3; \theta, \mathbf{A}, \pi)} = \prod_{p=2}^K \prod_{i=1}^{\mathbf{m}_{p,N}} \left[(1 - v(\mathcal{S}_{pi}, \mathbf{b}^3)) + v(\mathcal{S}_{pi}, \mathbf{b}^3) \prod_{ij \in \bar{\mathcal{S}}_{pi}} b(q_{ij}; \theta) \right],$$

where we recall the definition $\bar{\mathcal{S}}_{pi} = \mathcal{S}_{pi} \cup \mathcal{S}_{pi}^{fd} \cup \mathcal{S}_{pi,o}^{pd}$ with $\mathcal{S}_{pi}^{fd} = \{ij \in \mathcal{D}_p^{fd} | (i, j) \in v(\mathcal{S}_{pi}) \times v(\mathcal{S}_{pi})\}$ and $\mathcal{S}_{pi,o}^{pd} = \{ij \in \mathcal{D}_{p,o}^{pd} | (i, j) \in v(\mathcal{S}_{pi}) \times \mathcal{V}^o\}$.

Summing (26) over all $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ thus yields

$$\frac{\sum_{\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)} \Pr(\mathbf{D}^3 = \mathbf{b}^3; \theta, \mathbf{A}, \pi)}{\Pr(\mathbf{D}^3 = \mathbf{d}^3; \theta, \mathbf{A}, \pi)} = \prod_{p=2}^K \prod_{\mathbf{i}=1}^{\mathbf{m}_{p,N}} \left[1 + \prod_{ij \in \tilde{S}_{p\mathbf{i}}} b(q_{ij}; \theta) \right].$$

Note that the product to the right of the equality above evaluates to a sum of $2^{\sum_{p=2}^K \mathbf{m}_{p,N}}$ terms, one for each element of $\mathbb{B}_K(\mathbf{d}^3)$, as required. Each identifying star system $\mathbf{i} = \mathbf{1}, \dots, \mathbf{m}_{p,N}$ (for $p = 2, \dots, K$) either contributes 1 (when not swapped in $\mathbf{b}^3$) or $\prod_{ij \in \tilde{S}_{p\mathbf{i}}} b(q_{ij}; \theta)$ (when swapped in $\mathbf{b}^3$) to the product which forms each summand to the left of the equality.

Inverting yields (19) as claimed.

B.3 Proof of Theorem 2 (Null Sampler)

Denote the state transition graph of the Markov Chain defined by Algorithm 1 by $\Phi = (\mathcal{V}_\phi, \mathcal{A}_\phi)$. The vertex set, $\mathcal{V}_\phi$, coincides with the set of all network sequences in $\mathbb{B}_K(\mathbf{d}^3)$. We use v_ϕ to denote a generic element of the vertex set. We first characterize the arc set of the state transition graph, allowing us to verify irreducibility and ergodicity of the Markov chain. The main result then follows by checking conditions (i) and (ii) of Theorem 5.

The arc set, $\mathcal{A}_\phi$, is defined as follows.

1. For each vertex $v_\phi \in \mathcal{V}_\phi$ there exists the self-loop (v_ϕ, v_ϕ) with probability weight ρ (see Step 2 of Algorithm 1).
2. We say that v'_ϕ is a *neighbor* of v_ϕ (with $v_\phi \neq v'_\phi$) if it is possible to move (from network sequence) v_ϕ to v'_ϕ in a single step of our algorithm (i.e., by swapping the edges of a single star system). Let $\mathcal{N}(v_\phi)$ denote the set of vertices adjacent to v_ϕ . For each such neighbor, the arc (v_ϕ, v'_ϕ) is present in $\mathcal{A}_\phi$. Note $|\mathcal{N}(v_\phi)| = \mathbf{m}_{2:K,N}$ for all $v_\phi \in \mathcal{V}_\phi$. This follows since Algorithm 1, in step 3, selects exactly one out of the $\mathbf{m}_{2:K,N} = \sum_{p=2}^K \mathbf{m}_{p,N}$ stable star systems uniformly at random. Let $\tilde{\mathbf{b}}^3$ be the network sequence associated with vertex v'_ϕ and $\mathbf{b}^3$ the one associated with v_ϕ . The probability weight on arc (v_ϕ, v'_ϕ) equals

$$p(\mathbf{b}^3, \tilde{\mathbf{b}}^3) = \frac{1 - \rho}{\mathbf{m}_{2:K,N}} \min \left\{ 1, \frac{\ell^{css}(\tilde{\mathbf{b}}^3; \theta_0)}{\ell^{css}(\mathbf{b}^3; \theta_0)} \right\}. \quad (33)$$

3. Finally there exists an additional self loop (v_ϕ, v_ϕ) if all the arcs leaving v_ϕ , defined in bullet points 1 and 2 above do not have probability weights which sum to 1. The

probability attached to this additional self-loop is one minus the probability of all other arcs with source (tail) v_ϕ .

Let v_ϕ be the (unique) vertex associated with network sequence $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ and v'_ϕ the one associated with $\tilde{\mathbf{b}}^3 \in \mathbb{B}_K(\mathbf{d}^3)$. Let $\delta(v_\phi) = \ell^{css}(\mathbf{b}^3; \theta_0)$ and $\delta(v'_\phi) = \ell^{css}(\tilde{\mathbf{b}}^3; \theta_0)$ be the associated conditional star system likelihood probabilities. Let $L = 2^{\mathbf{m}_{2:K,N}}$ denote the cardinality of the set $\mathbb{B}_K(\mathbf{d}^3)$ and hence the number of vertices in the state transition graph. The $L \times L$ transition matrix of edge weights for this graph, $\mathbf{P} = [p_{v_\phi, v'_\phi}]_{v_\phi, v'_\phi \in \mathcal{V}_\phi}$, has elements

$$p_{v_\phi, v'_\phi} = \begin{cases} \frac{1-\rho}{\mathbf{m}_{2:K,N}} \min \left\{ 1, \frac{\delta(v'_\phi)}{\delta(v_\phi)} \right\} & \text{if } v_\phi \neq v'_\phi \text{ and } v'_\phi \in \mathcal{N}(v_\phi) \\ 0 & \text{if } v_\phi \neq v'_\phi \text{ and } v'_\phi \notin \mathcal{N}(v_\phi) \\ 1 - \sum_{v'_\phi \neq v_\phi} p_{v_\phi, v'_\phi} & \text{if } v_\phi = v'_\phi \end{cases} \quad (34)$$

Next observe that the state transition graph is strongly connected. Studying equation (33) we see that $p(\mathbf{b}^3, \tilde{\mathbf{b}}^3) > 0$ for all $\mathbf{b}^3, \tilde{\mathbf{b}}^3 \in \mathbb{B}_K(\mathbf{d}^3)$, and hence that the weight on arc (v_ϕ, v'_ϕ) is strictly positive for all $v'_\phi \in \mathcal{N}(v_\phi)$. Similarly for any such v'_ϕ we have $v_\phi \in \mathcal{N}(v'_\phi)$ and hence (v'_ϕ, v_ϕ) also endowed with positive weight. Since any $\tilde{\mathbf{b}}^3 \in \mathbb{B}_K(\mathbf{d}^3)$ can be constructed by swapping the edges of a finite number of stable star systems starting from any other $\mathbf{b}^3 \in \mathbb{B}_K(\mathbf{d}^3)$, there exists a directed path, with positive probability weight, between any two vertices in Φ . The state transition graph is strongly connected and hence the Markov chain defined by Algorithm 1 is irreducible. Due to the presence of self-loops, the state transition graph is not bipartite and therefore aperiodic. Since the chain is finite, irreducible, and aperiodic it is ergodic with a unique stationary distribution. We can use Theorem 5 to verify that this stationary distribution coincides with the claimed (null) sampling distribution.

To use the Theorem we associate $\pi_l = \delta(v_\phi)$, $\pi_m = \delta(v'_\phi)$, $p_{lm} = \frac{1-\rho}{\mathbf{m}_{2:K,N}} \min \{1, \delta(v'_\phi) / \delta(v_\phi)\}$ and $p_{ml} = \frac{1-\rho}{\mathbf{m}_{2:K,N}} \min \{1, \delta(v_\phi) / \delta(v'_\phi)\}$. Condition (i) of Theorem 5 is satisfied since $\ell^{css}(\cdot; \theta_0)$ is a probability mass function with support $\mathbb{B}_K(\mathbf{d}^3)$ and hence $\sum_{v_\phi \in \mathcal{V}_\phi} \delta(v_\phi) = 1$. Requirement (ii) is the so called *detailed balanced* condition.

To verify this condition let v_ϕ and v'_ϕ be two vertices in $\mathcal{V}_\phi$ labeled, without loss of generality, such that $\delta(v_\phi) \geq \delta(v'_\phi)$. If $v'_\phi \notin \mathcal{N}(v_\phi)$ or $v'_\phi = v_\phi$, then condition (ii) of Theorem 5 holds

trivially, otherwise for $v'_\phi \in \mathcal{N}(v_\phi)$

$$\begin{aligned}
\pi_l p_{lm} &= \delta(v_\phi) \frac{1-\rho}{\mathbf{m}_{2:K,N}} \min \left\{ 1, \frac{\delta(v'_\phi)}{\delta(v_\phi)} \right\} \\
&= \delta(v_\phi) \frac{1-\rho}{\mathbf{m}_{2:K,N}} \frac{\delta(v'_\phi)}{\delta(v_\phi)} \\
&= \delta(v'_\phi) \frac{1-\rho}{\mathbf{m}_{2:K,N}} \\
&= \delta(v'_\phi) \frac{1-\rho}{\mathbf{m}_{2:K,N}} \min \left\{ 1, \frac{\delta(v_\phi)}{\delta(v'_\phi)} \right\} \\
&= \pi_m p_{ml}
\end{aligned}$$

as required. The second equality follows from $\min \{1, \delta(v'_\phi) / \delta(v_\phi)\} = \delta(v'_\phi) / \delta(v_\phi)$ and the fourth from $\min \{1, \delta(v_\phi) / \delta(v'_\phi)\} = 1$ (both implications of $\delta(v_\phi) \geq \delta(v'_\phi)$). This gives a stationary distribution for our Markov chain of

$$\begin{aligned}
\pi &= (\delta(v_{\phi 1}), \dots, \delta(v_{\phi L}))' \\
&= (\ell^{css}(\mathbf{b}_1^3; \theta_0), \dots, \ell^{css}(\mathbf{b}_L^3; \theta_0))'
\end{aligned}$$

with $\mathbb{B}_K(\mathbf{d}^3) = \{\mathbf{b}_1^3, \dots, \mathbf{b}_L^3\}$ and $\mathcal{V}_\phi = \{v_{\phi 1}, \dots, v_{\phi L}\}$ (conformably arranged).

B.4 Proof of Theorem 3 (Consistency)

Define:

$$\begin{aligned}
\tilde{Q}_N(\theta) &:= \mathbb{E} \left[\hat{Q}_N(\theta) \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \\
&= \frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p=2}^K \sum_{i=\{i_1, i_2, \dots, i_p\}} Z_{pi} \cdot \mathbb{E} \left[\log \left(\frac{\exp(T'_{pi}\theta)}{1 + \exp(T'_{pi}\theta)} \right) \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right],
\end{aligned}$$

and recall the definition of $\hat{Q}_N(\theta)$ given in the main text. The proof involves verifying that the conditions of Lemma 5 in Appendix A are satisfied whenever Assumptions 1 to 4 hold. This involves the following steps:

1. Show uniform convergence to the target: $\sup_{\theta \in \Theta} |\hat{Q}_N(\theta) - \tilde{Q}_N(\theta)| \xrightarrow{P} 0$. This involves two sub-steps:

- (a) Demonstrating point-wise convergence, for all $\theta \in \Theta$: $\hat{Q}_N(\theta) - \tilde{Q}_N(\theta) \xrightarrow{P} 0$

(b) Inferring uniform convergence through a stochastic equicontinuity argument (following Newey (1991)).

2. Show that θ_0 is a well separated maximizer of $\tilde{Q}_N$, by establishing a lower bound on the hessian of $\tilde{Q}_N$.

Step 1.a: Fix θ ; by conditional independence (Lemma 4) we have

$$\begin{aligned} \mathbb{V}(\hat{Q}_N(\theta) \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}) \\ = \frac{1}{\left(\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]\right)^2} \sum_{p=2}^K \sum_{i=\{i_1, i_2, \dots, i_p\}} Z_{pi} \cdot \mathbb{V} \left(\ln \left(\frac{\exp(T'_{pi}\theta)}{1 + \exp(T'_{pi}\theta)} \right) \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right). \end{aligned}$$

Next observe that for all $x \in \mathbb{R}^1$:

$$\min\{0, x\} - \ln(2) \leq \ln \left(\frac{\exp(x)}{1 + \exp(x)} \right) \leq \min\{0, x\}, \quad (35)$$

and hence

$$\mathbb{V} \left(\ln \left(\frac{\exp(T'_{pi}\theta)}{1 + \exp(T'_{pi}\theta)} \right) \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right) \leq \mathbb{E} \left[(|T'_{pi}\theta| + \ln(2))^2 \right].$$

Therefore, by Assumptions 1 and 3:

$$\mathbb{V} \left(\hat{Q}_N(\theta) \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right) \leq \frac{1}{\left(\mathbb{E} \left[\sum_{i,p} Z_{i,p} \right]\right)^2} \sum_{p=2}^K \sum_{i=\{i_1, i_2, \dots, i_p\}} Z_{pi} \cdot \mathbb{E} \left[(|T'_{pi}\theta| + \ln(2))^2 \right],$$

and consequently

$$\mathbb{E} \left[\mathbb{V} \left(\hat{Q}_N(\theta) \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right) \right] \leq \frac{C}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]}$$

for some constant C (by Assumption 1 and 3).

Assumption 4, and non-negativity of variances, then yields $\mathbb{V} \left(\hat{Q}_N(\theta) \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right) \xrightarrow{p} 0$. From this we conclude that $\hat{Q}_N(\theta) - \mathbb{E} \left[\hat{Q}_N(\theta) \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \xrightarrow{p} 0$ (i.e., $\hat{Q}_N(\theta) - \tilde{Q}_N(\theta) \xrightarrow{p} 0$ for any θ). This completes Step 1.a.

Step 1.b: We demonstrate uniform convergence by verifying the conditions of Corollary 2.2 in Newey (1991) for $\Delta Q_N := \hat{Q}_N - \tilde{Q}_N$. Fix $\theta, \tilde{\theta} \in \Theta$, by Assumption 1 and the mean value

theorem, there exists some $\bar{\theta} \in (\theta, \tilde{\theta})$ such that

$$\begin{aligned} \Delta Q_N(\theta) - \Delta Q_N(\tilde{\theta}) &= \frac{\partial \Delta Q_N}{\partial \theta}(\bar{\theta})(\theta - \tilde{\theta}) \\ &= \frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \\ &\quad \times \sum_{p,i} Z_{pi} \cdot ((1 - \Lambda(T'_{pi}\bar{\theta}))T'_{pi} - \mathbb{E} [(1 - \Lambda(T'_{pi}\bar{\theta}))T'_{pi} | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}]) \\ &\quad \times (\theta - \tilde{\theta}), \end{aligned}$$

recalling that $\Lambda(x) \stackrel{def}{=} \frac{\exp(x)}{1+\exp(x)}$. By the TI and CSI we have

$$|\Delta Q_N(\theta) - \Delta Q_N(\tilde{\theta})| \leq \frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} \cdot (\|T_{pi}\| + \mathbb{E} [\|T_{pi}\| | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}]) \cdot \|\theta - \tilde{\theta}\|.$$

To apply Corollary 2.2 of Newey (1991), it suffices to show that

$$B_N := \frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} (\|T_{pi}\| + \mathbb{E} [\|T_{pi}\| | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}]) = O_p(1)$$

which is established by noting that $\mathbb{E}[B_N]$ is bounded uniformly in N by Assumption 3.²⁴ We therefore conclude that

$$\sup_{\theta \in \Theta} |\hat{Q}_N(\theta) - \tilde{Q}_N(\theta)| \xrightarrow{p} 0$$

by applying Corollary 2.2 in Newey (1991), where we take Newey's $\hat{Q}_N$ to be ΔQ_N in this proof and Newey's $\tilde{Q}_N$ to be the constant 0 function. This completes Step 1.b.

Step 2: First, we show that $\tilde{Q}_N$ is maximized at θ_0 for all N . The conditional likelihood of $\mathbf{d}^3$ given $\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}_p, \mathbf{A}$ is²⁵

$$\Pr(\mathbf{d}^3 | \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}_p, \mathbf{A}) = \prod_{\mathbf{i}} \left(\frac{\exp(T'_{pi}\theta_0)}{1 + \exp(T'_{pi}\theta_0)} \right)^{Z_{pi}}$$

²⁴Note that Assumption 3 is expressed in terms of moments of T_{pi} conditional on $(\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z})$, not $(\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A})$. A consequence of Lemma 4 is that the distribution of T_{pi} conditional on $\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A}$ is independent of $\mathbf{A}$. Hence, the moments of T_{pi} under both sets of conditioning variables are equal.

²⁵Let $\mathbb{B}_p^+(\mathbf{d}^3)$ denote the set of network sequences reachable by swapping the edges of identifying star systems of *size p alone*. The information content of the event $\mathbf{D}^3 \in \mathbb{B}_p^+(\mathbf{d}^3)$, and $\mathbf{Z}_p$ taking a particular configuration, coincides.

therefore, for almost any $\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}_p, \mathbf{A}$, the function $\theta \rightarrow \mathbb{E} \left[\ln \left(\prod_{\mathbf{i}} \left(\frac{\exp(T'_{p\mathbf{i}}\theta)}{1+\exp(T'_{p\mathbf{i}}\theta)} \right)^{Z_{p\mathbf{i}}} \right) \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}_p, \mathbf{A} \right]$ is maximized at θ_0 .

Fix $\epsilon > 0$. Concavity of $\tilde{Q}_N$ in θ and being maximized at θ_0 for every N , implies that $\sup_{\|\theta - \theta_0\| \geq \epsilon} \tilde{Q}_N(\theta)$ is achieved at some θ_ϵ such that $\|\theta_0 - \theta_\epsilon\| = \epsilon$. The multivariate mean value theorem (and by Assumption 1) gives, for some $\bar{\theta}_\epsilon \in (\theta, \theta_\epsilon)$ and defining $\Delta\theta := \theta_\epsilon - \theta_0$:

$$\begin{aligned} \tilde{Q}_N(\theta_\epsilon) - \tilde{Q}_N(\theta_0) &= \frac{\partial \tilde{Q}_N}{\partial \theta}(\theta_0) \Delta\theta + \Delta\theta' \frac{\partial^2 \tilde{Q}_N}{\partial \theta^2}(\bar{\theta}_\epsilon) \Delta\theta \\ &= 0 + \Delta\theta' \frac{\partial^2 \tilde{Q}_N}{\partial \theta^2}(\bar{\theta}_\epsilon) \Delta\theta \\ &= - \frac{1}{\mathbb{E} \left(\sum_{p\mathbf{i}} Z_{p\mathbf{i}} \right)} \\ &\quad \times \sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \cdot \Delta\theta' \mathbb{E} \left[\Lambda(T'_{p\mathbf{i}}\bar{\theta}_\epsilon)(1 - \Lambda(T'_{p\mathbf{i}}\bar{\theta}_\epsilon)) T_{p\mathbf{i}} T'_{p\mathbf{i}} \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \Delta\theta. \end{aligned}$$

Fix some constant $a > 0$; for any $\mathbf{i}$ and p we have, using the MI and CSI,

$$\begin{aligned} \Delta\theta' \mathbb{E} \left[\Lambda(T'_{p\mathbf{i}}\bar{\theta}_\epsilon)(1 - \Lambda(T'_{p\mathbf{i}}\bar{\theta}_\epsilon)) T_{p\mathbf{i}} T'_{p\mathbf{i}} \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \Delta\theta &\geq \Lambda(a)(1 - \Lambda(a)) \\ &\quad \times \mathbb{E} \left[\Delta\theta' T_{p\mathbf{i}} T'_{p\mathbf{i}} \Delta\theta \mathbf{1}(|T'_{p\mathbf{i}}\bar{\theta}_\epsilon| \leq a) \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \\ &\geq \Lambda(a)(1 - \Lambda(a)) \left\{ \mathbb{E} \left[\Delta\theta' T_{p\mathbf{i}} T'_{p\mathbf{i}} \Delta\theta \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right. \\ &\quad \left. - \mathbb{E} \left[\Delta\theta' T_{p\mathbf{i}} T'_{p\mathbf{i}} \Delta\theta \mathbf{1}(|T'_{p\mathbf{i}}\bar{\theta}_\epsilon| \geq a) \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right\} \\ &\geq \Lambda(a)(1 - \Lambda(a)) \left\{ \mathbb{E} \left[\Delta\theta' T_{p\mathbf{i}} T'_{p\mathbf{i}} \Delta\theta \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right. \\ &\quad \left. - \frac{1}{a} \left(\mathbb{E} \left[(\Delta\theta' T_{p\mathbf{i}} T'_{p\mathbf{i}} \Delta\theta)^2 \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \right. \\ &\quad \left. \times \left(\mathbb{E} \left[\bar{\theta}'_\epsilon T_{p\mathbf{i}} T'_{p\mathbf{i}} \bar{\theta}_\epsilon \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \right\}. \end{aligned}$$

Combining the results above we get:

$$\begin{aligned}
\tilde{Q}_N(\theta_0) - \tilde{Q}_N(\theta_\epsilon) &\geq \frac{\Lambda(a)(1 - \Lambda(a))}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} \cdot \left\{ \mathbb{E} \left[\Delta \theta' T_{pi} T_{pi}' \Delta \theta \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right. \\
&\quad - \frac{1}{a} \left(\mathbb{E} \left[(\Delta \theta' T_{pi} T_{pi}' \Delta \theta)^2 \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[\bar{\theta}'_\epsilon T_{pi} T_{pi}' \bar{\theta}_\epsilon \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \Big\} \\
&\geq \Lambda(a)(1 - \Lambda(a)) \left\{ \Delta \theta' \mathbb{E} \left[\frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} T_{pi} T_{pi}' \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \Delta \theta \right. \\
&\quad - \frac{1}{a} \left(\frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} \mathbb{E} \left[(\Delta \theta' T_{pi} T_{pi}' \Delta \theta)^2 \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \\
&\quad \times \left(\frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} \mathbb{E} \left[\bar{\theta}'_\epsilon T_{pi} T_{pi}' \bar{\theta}_\epsilon \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \Big\} \\
&\geq \Lambda(a)(1 - \Lambda(a)) \left\{ \epsilon^2 C_1 \right. \\
&\quad - \frac{1}{a} \left(\frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} \|\Delta \theta\|^4 \mathbb{E} \left[\|T_{pi}\|^4 \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \\
&\quad \times \left(\sup_{\Theta} \|\theta\|^2 \frac{1}{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \sum_{p,i} Z_{pi} \mathbb{E} \left[\|T_{pi}\|^2 \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right)^{1/2} \Big\} \\
&\geq \Lambda(a)(1 - \Lambda(a)) \left[\epsilon^2 C_1 - \frac{1}{a} \sqrt{\epsilon^4 C_2} \sqrt{\sup_{\Theta} \|\theta\|^2 \sqrt{C_2}} \right], \tag{36}
\end{aligned}$$

where the first relationship follows immediately from the calculations above, the second from Jensen's inequality, the third from Assumption 1, Assumption 2, and the relationship $\text{tr}(ba'ab')^{1/2} = (b'ba'a)^{1/2} = \|a\| \|b\|$, and – finally – the fourth from Assumptions 2 and 3.

Finally, if we choose an $a > 0$ such that

$$\delta(\epsilon) := \Lambda(a)(1 - \Lambda(a)) \left[C_1 \epsilon^2 - \frac{1}{a} \epsilon^2 \sqrt{\sup_{\Theta} \|\theta\|^2 C_2^{\frac{3}{4}}} \right] > 0,$$

we have

$$\left| \tilde{Q}_N(\theta_0) - \sup_{\|\theta - \theta_0\| \geq \epsilon} \tilde{Q}_N(\theta) \right| \geq \delta(\epsilon)$$

as required. This completes step 2.

All the conditions of Lemma 5 are met, so we can conclude: $\hat{\theta} \rightarrow_p \theta_0$.

B.5 Proof of Theorem 4 (Asymptotic Normality)

Define

$$L_N(\theta) := \frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \sum_{p=2}^K \sum_i l_{pi}(\theta)$$

with

$$l_{pi}(\theta) := Z_{pi} \log \left(\frac{\exp(T'_{pi}\theta)}{1 + \exp(T'_{pi}\theta)} \right)$$

A mean value expansion in θ around θ_0 , in conjunction with Assumption 1, yields some $\bar{\theta} \in (\hat{\theta}, \theta_0)$ such that

$$\begin{aligned} 0 &= \frac{\partial L_N}{\partial \theta}(\hat{\theta}) = \frac{\partial L_N}{\partial \theta}(\theta_0) + \frac{\partial^2 L_N}{\partial \theta^2}(\theta_0)(\hat{\theta} - \theta_0) + \sum_k (\hat{\theta}_k - \theta_{0,k}) \frac{\partial^3 L_N}{\partial \theta_k \partial \theta^2}(\bar{\theta})(\hat{\theta} - \theta_0) \\ &= \frac{\partial L_N}{\partial \theta}(\theta_0) + \frac{\partial^2 L_N}{\partial \theta^2}(\theta_0)(\hat{\theta} - \theta_0) + o_p(\hat{\theta} - \theta_0), \end{aligned}$$

since

$$\frac{\partial^3 L_N}{\partial \theta_j \partial \theta_k \partial \theta_l}(\bar{\theta}) = - \frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \sum_{p,i} Z_{pi} \cdot \Lambda(T'_{pi}\bar{\theta}) (1 - \Lambda(T'_{pi}\bar{\theta})) T_{pi,j} T_{pi,k} T_{pi,l},$$

and by Assumption 3, the TI, the LIE, and JI (observing that $f(x) = x^{3/4}$ is concave):

$$\begin{aligned} \mathbb{E} \left[\left\| \frac{\partial^3 L_N}{\partial \theta_j \partial \theta_k \partial \theta_l}(\bar{\theta}) \right\| \right] &\leq \frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \sum_{p,i} \mathbb{E} [|Z_{pi} \cdot T_{pi,j} T_{pi,k} T_{pi,l}|] \\ &\leq \frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \sum_{p,i} \mathbb{E} [Z_{pi} \cdot \|T_{pi}\|^3] \\ &\leq \frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \mathbb{E} \left[\left(\sum_{p,i} Z_{pi} \right) \mathbb{E} \left[\frac{\sum_{p,i} Z_{pi} \|T_{pi}\|^3}{\sum_{p,i} Z_{pi}} \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right] \\ &\leq \frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \mathbb{E} \left[\left(\sum_{p,i} Z_{pi} \right) \mathbb{E} \left[\frac{\sum_{p,i} Z_{pi} \|T_{pi}\|^4}{\sum_{p,i} Z_{pi}} \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right]^{3/4} \right] \\ &\leq \frac{1}{\sum_{p,i} \mathbb{E}[Z_{pi}]} \mathbb{E} \left[\left(\sum_{p,i} Z_{pi} \right) (C_2)^{3/4} \right] \leq C_2^{3/4}, \end{aligned}$$

where the last inequality is a direct implication of Assumption 3. We consequently have

$$\hat{\theta} - \theta_0 = - \left(\frac{\partial^2 L_N}{\partial \theta^2}(\theta_0) + o_p(1) \right)^+ \frac{\partial L_N}{\partial \theta}(\theta_0)$$

where a ‘+’ superscript denotes the generalized inverse. To demonstrate asymptotic normality of $\hat{\theta}$ we therefore need to show that (i) $\frac{\partial L_N}{\partial \theta}(\theta_0)$, suitably normalized, asymptotically behaves like a normal random variable and (ii) $\frac{\partial^2 L_N}{\partial \theta^2}(\theta_0)$ converges in probability to a positive definite matrix.

Asymptotic normality of $\frac{\partial L_N}{\partial \theta}(\theta_0)$: We use the Wald Device. Fix some $\lambda \in \mathbb{R}^2$. Note that, by conditional independence:

$$\mathbb{V} \left(\sum_{p, \mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0) \right) = \sum_{p, \mathbf{i}} \mathbb{V} (\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0)) = \sum_{p, \mathbf{i}} \mathbb{E} \left[(\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2 \right].$$

Define the algebra: $\mathcal{G}_N := \sigma(\mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A})$, and enumerate the tuples $(p, \mathbf{i})$ so that every tuple $(p, \mathbf{i})$ corresponds to an integer $i \in \{1, \dots\}$. Define the filtration $\mathcal{F}_{N,i} := \sigma(\mathcal{G}_N, t_1, t_2, \dots, t_i)$. Define $W_i := \frac{1}{\sqrt{\mathbb{V}(\sum_{p, \mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))}} \lambda' \nabla_{\theta} l_i(\theta_0)$. We verify the conditions of Theorem 2.3 in McLeish (1974) for the martingale difference W_i :

- (a) $\mathbb{E} [\max_i |W_i|^2] \leq \sum_i \mathbb{E} [|W_i|^2] = 1$, so $\max_i |W_i|$ is uniformly bounded in L_2 .
- (b) Fix $\epsilon > 0$. Note that $\Pr(\max_i |W_i| > \epsilon) \leq \Pr(\sum_i W_i^2 1(|W_i| > \epsilon) > \epsilon^2)$.²⁶ Moreover:

$$\sum_i W_i^2 \cdot 1(|W_i| > \epsilon) \leq \frac{1}{\epsilon^2} \sum_i W_i^4 = \frac{1}{\epsilon^2} \frac{\sum_{p, \mathbf{i}} (\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^4}{\left(\sum_{p, \mathbf{i}} \mathbb{E} [(\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2] \right)^2}.$$

²⁶This follows because occurrence of the event on the right implies that the one on the left also occurs.

Next observe that, using Assumption 3:

$$\begin{aligned}
\mathbb{E} \left[\sum_{p,\mathbf{i}} (\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^4 \right] &= \sum_{p,\mathbf{i}} \mathbb{E} \left[(\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^4 \right] \\
&= \sum_{p,\mathbf{i}} \mathbb{E} \left[(1 - \Lambda(T'_{p\mathbf{i}} \theta_0))^4 \cdot Z_{p\mathbf{i}} \cdot (\lambda' T_{p\mathbf{i}})^4 \right] \\
&\leq \|\lambda\|^4 \sum_{p,\mathbf{i}} \mathbb{E} [\|T_{p\mathbf{i}}\|^4 Z_{p\mathbf{i}}] \\
&\leq \|\lambda\|^4 \cdot \mathbb{E} \left[\left(\sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \right) \mathbb{E} \left[\frac{\sum_{p,\mathbf{i}} \|T_{p\mathbf{i}}\|^4 \cdot Z_{p\mathbf{i}}}{\sum_{p,\mathbf{i}} Z_{p\mathbf{i}}} \middle| \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] \right] \\
&\leq C_2 \cdot \|\lambda\|^4 \cdot \mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \right].
\end{aligned}$$

Moreover, via an argument similar to that used to derive (36) above, there exists a constant C_1 such that

$$\sum_{p,\mathbf{i}} \mathbb{E} [(\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2] \geq C_1 \cdot \|\lambda\|^2 \cdot \mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \right],$$

and hence we have that:

$$\mathbb{E} \left[\sum_i W_i^2 \cdot 1(|W_i| > \epsilon) \right] \leq \frac{1}{\epsilon^2} \frac{C_2 \cdot \|\lambda\|^4 \cdot \sum_{p,\mathbf{i}} \mathbb{E} [Z_{p\mathbf{i}}]}{C_1^2 \cdot \|\lambda\|^4 \cdot \left(\sum_{p,\mathbf{i}} \mathbb{E} [Z_{p\mathbf{i}}] \right)^2},$$

which implies, invoking Assumption 4, that:

$$\mathbb{E} \left[\sum_i W_i^2 \cdot 1(|W_i| > \epsilon) \right] \rightarrow 0, \tag{37}$$

and hence that $\max_i |W_i| \xrightarrow{p} 0$ as required.

- (c) This requirement is satisfied by Theorem 2.23 in Hall and Heyde (1980). Condition 2.24 in that reference follows from observing that:

$$\sum_{i=1}^n \mathbb{E} [W_i^2 | \mathcal{F}_{n,i-1}] = \frac{\mathbb{V} \left(\sum_{p,\mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0) \middle| \mathcal{G}_N \right)}{\mathbb{V} \left(\sum_{p,\mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0) \right)},$$

then, by the MI, for any $M > 0$

$$\begin{aligned} \Pr \left(\sum_{i=1}^n \mathbb{E} [W_i^2 | \mathcal{F}_{n,i-1}] > M \right) &\leq \frac{\mathbb{E} \left(\mathbb{V} \left(\sum_{p,i} \lambda' \nabla_{\theta} l_{pi}(\theta_0) \middle| \mathcal{G}_N \right) \right)}{M \times \mathbb{V} \left(\sum_{p,i} \lambda' \nabla_{\theta} l_{pi}(\theta_0) \right)} \\ &\leq \frac{1}{M} \end{aligned}$$

where the second inequality follows by the law of total variance. So that:

$$\sup_N \Pr \left(\sum_{i=1}^n \mathbb{E} [W_i^2 | \mathcal{F}_{n,i-1}] > M \right) \rightarrow 0; \text{ as } M \rightarrow +\infty$$

as desired and required. Condition 2.25 of Hall and Heyde (1980) (the conditional Lindeberg-Feller condition) follows from equation (37) and an application of the MI. Fix $M > 0$ and $\epsilon > 0$:

$$\begin{aligned} \Pr \left(\sum_i \mathbb{E} [W_i^2 \cdot 1(|W_i| > \epsilon) | \mathcal{F}_{N,i-1}] \geq M \right) &\leq \frac{\mathbb{E} [\sum_i \mathbb{E} [W_i^2 \cdot 1(|W_i| > \epsilon) | \mathcal{F}_{N,i-1}]]}{M} \\ &= \frac{\mathbb{E} [\sum_i W_i^2 \cdot 1(|W_i| > \epsilon)]}{M} \\ &\rightarrow_{N \rightarrow +\infty} 0 \end{aligned}$$

Then Equation 2.26 in Hall and Heyde (1980) shows that $\sum W_i^2 \rightarrow_p 1$.

Steps (a), (b) and (c) above allow us to conclude that:

$$\frac{1}{\sqrt{\mathbb{V} \left(\sum_{p,i} \lambda' \nabla_{\theta} l_{pi}(\theta_0) \right)}} \sum_i \lambda' \nabla_{\theta} l_i(\theta_0) \rightarrow_d \mathcal{N}(0, 1)$$

By Assumption 5, Slutsky's lemma, as well as the information matrix inequality, we then get:

$$\frac{1}{\sqrt{\lambda' \mathcal{I}(\theta_0) \lambda}} \cdot \frac{1}{\sqrt{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]}} \cdot \sum_i \lambda' \nabla_{\theta} l_i(\theta_0) \rightarrow_d \mathcal{N}(0, 1).$$

Since λ is arbitrary, we have:

$$\sqrt{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \cdot \frac{\partial L_N(\theta_0)}{\partial \theta} \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta_0)). \quad (38)$$

Convergence of $\frac{\partial^2 L_N}{\partial \theta^2}(\theta_0)$: Notice that, for any λ , by the information matrix equality, almost surely for every $p, \mathbf{i}$:

$$\mathbb{E} \left[(\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2 + \lambda' \nabla_{\theta\theta} l_{p\mathbf{i}}(\theta_0) \lambda \mid \mathbf{D}_0, \mathbf{D}_3, \mathbf{Z}, \mathbf{A} \right] = 0,$$

and hence we get, using the conditional mean zero property of the score vector and conditional independence (i.e., Lemma 4),

$$\mathbb{V} \left(\frac{\sum_{p,\mathbf{i}} (\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2 + \lambda' \nabla_{\theta\theta} l_{p\mathbf{i}}(\theta_0) \lambda}{\mathbb{V} \left(\sum_{p,\mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0) \right)} \right) = \frac{\sum_{p,\mathbf{i}} \mathbb{E} \left[((\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2 + \lambda' \nabla_{\theta\theta} l_{p\mathbf{i}}(\theta_0) \lambda)^2 \right]}{\left(\sum_{p,\mathbf{i}} \mathbb{E} [(\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2] \right)^2}.$$

By arguments similar to those used to derive equation (36), we can show that $\sum_{p,\mathbf{i}} \mathbb{E} [(\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2]$ is lower bounded by a constant times $\mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \right]$ and that $\sum_{p,\mathbf{i}} \mathbb{E} \left[((\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2 + \lambda' \nabla_{\theta\theta} l_{p\mathbf{i}}(\theta_0) \lambda)^2 \right]$ is upper bounded by another constant times $\mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \right]$, then by Assumption 4:

$$\mathbb{V} \left(\frac{\sum_{p,\mathbf{i}} (\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2 + \lambda' \nabla_{\theta\theta} l_{p\mathbf{i}}(\theta_0) \lambda}{\mathbb{V} \left(\sum_{p,\mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0) \right)} \right) \rightarrow 0,$$

By our demonstration of pointwise convergence of the conditional star system log-likelihood in Step 1.a of Theorem 3 above, $\frac{\sum_{p,\mathbf{i}} (\lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0))^2}{\mathbb{V} \left(\sum_{p,\mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0) \right)} \xrightarrow{p} 1$ and hence, again invoking the information matrix equality,

$$\frac{-1}{\mathbb{V} \left(\sum_{p,\mathbf{i}} \lambda' \nabla_{\theta} l_{p\mathbf{i}}(\theta_0) \right)} \sum_{p,\mathbf{i}} \lambda' \nabla_{\theta\theta} l_{p\mathbf{i}}(\theta_0) \lambda \xrightarrow{p} 1.$$

Assumption 5 then implies:

$$\frac{1}{\mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \right]} \sum_{p,\mathbf{i}} \lambda' \nabla_{\theta\theta} l_{p\mathbf{i}}(\theta_0) \lambda \xrightarrow{p} -\lambda' \mathcal{I}(\theta_0) \lambda.$$

Since convergence happens for any λ , we therefore have:²⁷

$$\frac{1}{\mathbb{E} \left[\sum_{p,\mathbf{i}} Z_{p\mathbf{i}} \right]} \frac{\partial^2 L_N(\theta_0)}{\partial \theta^2} \xrightarrow{p} -\mathcal{I}(\theta_0). \quad (39)$$

²⁷We use the identity: $x'Hy = \frac{1}{4} \left[(x+y)'H(x+y) - (x-y)'H(x-y) \right]$ for any symmetric H and any vectors x and y .

Equation (38) and (39) allow to conclude that:

$$\sqrt{\mathbb{E} \left[\sum_{p,i} Z_{pi} \right]} \cdot (\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}(0, \mathcal{I}^{-1}(\theta_0)),$$

as claimed.

B.6 Proof of Proposition 2

We consider the two cases $\beta_0 \geq 0$ and $\beta_0 < 0$ separately.

Case 1: $\beta_0 \geq 0$

We show that for any dyad $i \neq j$, the probability of swapping an edge between periods 1 and 2 decays exponentially with N . To start, we present and prove a Lemma which controls the tails of R_{ij0} and R_{ij1} .

Lemma 6. *Assume $\beta_0 \geq 0$. Define $c_0 := \frac{1}{2}p^2$ and $c_1 := \frac{1}{2}\Lambda(-|\alpha_0| - \bar{A})^2$, there exist $\gamma_0 > 0$ and $\gamma_1 > 0$ independent of N such that:*

$$\mathbb{P}(\exists i, j : R_{ij0} \leq c_0 N) \leq N^2 \exp(-\gamma_0 N) \text{ and } \mathbb{P}(\exists i, j : R_{ij1} \leq c_1 N) \leq N^2 \exp(-\gamma_1 N).$$

Proof of Lemma 6. We start by observing that $R_{ij0} = \sum_k D_{ik0} D_{jk0}$ is a sum of $N - 2$ independent Bernoulli random variables each with success probability p^2 and hence $\mathbb{E}[R_{ij0}] = (N - 2)p^2$. Hoeffding's Inequality (HI) then gives, for N large enough:

$$\begin{aligned} \mathbb{P}(\exists i, j : R_{ij0} \leq c_0 N) &\leq \sum_{i < j} \mathbb{P}(R_{ij0} \leq c_0 N) \\ &\leq \sum_{i < j} \mathbb{P}(R_{ij0} - \mathbb{E}[R_{ij0}] \leq c_0 N - \mathbb{E}[R_{ij0}]) \\ &\leq \binom{N}{2} \times \exp\left(-\frac{2[(N - 2)p^2 - c_0 N]^2}{N - 2}\right) \\ &\leq N^2 \exp\left(-\frac{Np^4}{6}\right). \end{aligned}$$

So the first hypothesis holds with $\gamma_0 = \frac{p^4}{6}$. Next consider the second hypothesis. By the law

of total probability:

$$\begin{aligned}\mathbb{P}(\exists i, j : R_{ij1} \leq c_1 N) &\leq \sum_{i < j} \mathbb{P}(R_{ij1} \leq c_1 N) \\ &\leq \sum_{i < j} \mathbb{E} \left[\mathbb{P} \left(R_{ij1} - \mathbb{E}[R_{ij1} | \mathbf{D}_0, \mathbf{A}] \leq c_1 N - \mathbb{E}[R_{ij1} | \mathbf{D}_0, \mathbf{A}] \mid \mathbf{D}_0, \mathbf{A} \right) \right].\end{aligned}$$

Recalling that $A_{ij} \in [-\bar{A}, \bar{A}]$ we get:

$$\begin{aligned}\mathbb{E}[R_{ij1} | \mathbf{D}_0, \mathbf{A}] &= \sum_{k \neq i, j} \Lambda(\alpha_0 D_{ik0} + \beta_0 R_{ik0} + A_{ik}) \Lambda(\alpha_0 D_{jk0} + \beta_0 R_{jk0} + A_{jk}) \\ &\geq \sum_{k \neq i, j} \Lambda(-|\alpha_0| + \beta_0 R_{ik0} - \bar{A}) \Lambda(-|\alpha_0| + \beta_0 R_{jk0} - \bar{A}) \\ &\geq (N-2) \Lambda(-|\alpha_0| - \bar{A})^2.\end{aligned}$$

Invoking the HI for a second time yields, for N large enough:

$$\begin{aligned}\mathbb{P} \left(R_{ij1} - \mathbb{E}[R_{ij1} | \mathbf{D}_0, \mathbf{A}] \leq c_1 N - \mathbb{E}[R_{ij1} | \mathbf{D}_0, \mathbf{A}] \mid \mathbf{D}_0, \mathbf{A} \right) &\leq \exp \left(-\frac{2 \left(\mathbb{E}[R_{ij1} | \mathbf{D}_0, \mathbf{A}] - c_1 N \right)^2}{N-2} \right) \\ &\leq \exp \left(-\frac{2 \left(\left[\frac{N-4}{2} \right] \Lambda(-|\alpha_0| - \bar{A})^2 \right)^2}{N-2} \right) \\ &\leq \exp \left(-\frac{N \Lambda(-|\alpha_0| - \bar{A})^4}{6} \right).\end{aligned}$$

Then

$$\mathbb{P}(\exists i, j : R_{ij1} \leq c_1 N | \mathbf{D}_0, \mathbf{A}) \leq N^2 \exp \left(-\frac{N \Lambda(-|\alpha_0| - \bar{A})^4}{6} \right)$$

So the second hypothesis holds with $\gamma_1 = \frac{\Lambda(-|\alpha_0| - \bar{A})^4}{6}$.

Returning to the proof of Proposition 2. Let

$$\mathcal{E}_0 = \{\forall k, l : R_{kl0} > c_0 N\} \text{ and } \mathcal{E}_1 = \{\forall k, l : R_{kl1} > c_1 N\}.$$

By Lemma 6, $\mathbb{P}(\mathcal{E}_0^c) \leq N^2 \exp(-\gamma_0 N)$ and $\mathbb{P}(\mathcal{E}_1^c) \leq N^2 \exp(-\gamma_1 N)$.

Fix a pair i, j . Conditional on $\mathbf{D}^1$ and $\mathbf{A}$, on the event $\mathcal{E}_1$, we have $R_{kl1} > c_1 N$ for all k, l .

Hence:

$$\begin{aligned}
\mathbb{P}(D_{ij1} \neq D_{ij2} | \mathbf{D}^1, \mathbf{A}) 1(\mathcal{E}_1) &\leq 1(D_{ij1} = 0) + \mathbb{P}(D_{ij2} = 0 | D_{ij1} = 1, \mathbf{D}^1, \mathbf{A}) 1(\mathcal{E}_1) \\
&\leq 1(D_{ij1} = 0) + (1 - \Lambda(\alpha_0 + \beta_0 c_1 N - \bar{A})) \\
&\leq 1(D_{ij1} = 0) + e^{-\alpha_0 + \bar{A}} e^{-\beta_0 c_1 N}.
\end{aligned}$$

The probability that an arbitrary K -system $i := \{i_1, \dots, i_K\}$ is identifying, conditional on $\mathbf{D}^1, \mathbf{A}$ and restricted to $\mathcal{E}_1$, is bounded above by:

$$\begin{aligned}
\mathbb{P}(Z_i = 1 | \mathbf{D}^1, \mathbf{A}) 1(\mathcal{E}_1) &\leq \prod_{l \neq 1} \mathbb{P}(D_{i_1 i_l 1} \neq D_{i_1 i_l 2} | \mathbf{D}^1, \mathbf{A}) 1(\mathcal{E}_1) \\
&\leq \prod_{l \neq 1} \left(1(D_{i_1 i_l 1} = 0) + e^{-\alpha_0 + \bar{A}} e^{-\beta_0 c_1 N} \right).
\end{aligned}$$

Since the period-1 links form independently conditional on $\mathbf{D}^0, \mathbf{A}$, taking the expectation over $\mathbf{D}^1$ gives:

$$\mathbb{E}[\mathbb{P}(Z_i = 1 | \mathbf{D}^1, \mathbf{A}) 1(\mathcal{E}_1) | \mathbf{D}^0, \mathbf{A}] \leq \prod_{l \neq 1} \left(\mathbb{P}(D_{i_1 i_l 1} = 0 | \mathbf{D}^0, \mathbf{A}) + e^{-\alpha_0 + \bar{A}} e^{-\beta_0 c_1 N} \right).$$

On the event $\mathcal{E}_0$, $R_{kl0} > c_0 N$ for all k, l , so $\mathbb{P}(D_{i_1 i_l 1} = 0 | \mathbf{D}^0, \mathbf{A}) \leq e^{|\alpha_0| + \bar{A}} e^{-\beta_0 c_0 N}$. Thus:

$$\begin{aligned}
\mathbb{P}(Z_i = 1, \mathcal{E}_1 | \mathbf{D}^0, \mathbf{A}) 1(\mathcal{E}_0) &\leq \prod_{l \neq 1} \left(e^{|\alpha_0| + \bar{A}} e^{-\beta_0 c_0 N} + e^{-\alpha_0 + \bar{A}} e^{-\beta_0 c_1 N} \right) \\
&\leq C^{K-1} \exp(-\gamma_3 (K-1)N),
\end{aligned}$$

for some constants $C > 0$ and $\gamma_3 > 0$ independent of N, K . Consequently, taking the unconditional expectation ensures:

$$\mathbb{P}(Z_i = 1, \mathcal{E}_0 \cap \mathcal{E}_1) \leq C^{K-1} \exp(-\gamma_3 (K-1)N).$$

To bound the expected number of identifying stars, we partition the expectation. By Corollary 1, any dyad belongs to at most one identifying star, thus the total number of identifying

stars is deterministically bounded by $N/2$. We then obtain:

$$\begin{aligned}
\mathbb{E} \left[\sum_K m_{K,N} \right] &= \mathbb{E} \left[\sum_K m_{K,N} 1(\mathcal{E}_0 \cap \mathcal{E}_1) \right] + \mathbb{E} \left[\sum_K m_{K,N} 1((\mathcal{E}_0 \cap \mathcal{E}_1)^c) \right] \\
&\leq N \sum_{K=2}^{N-1} \binom{N-1}{K-1} \cdot \mathbb{P}(Z_i = 1, \mathcal{E}_0 \cap \mathcal{E}_1) + \frac{N}{2} \mathbb{P}(\mathcal{E}_0^c \cup \mathcal{E}_1^c) \\
&\leq N \sum_{K=2}^{N-1} \binom{N-1}{K-1} (C \exp(-\gamma_3 N))^{K-1} + \frac{N}{2} (N^2 \exp(-\gamma_0 N) + N^2 \exp(-\gamma_1 N)) \\
&= N \left([1 + C \exp(-\gamma_3 N)]^{N-1} - 1 - [C \exp(-\gamma_3 N)]^{N-1} \right) \\
&\quad + \frac{N^3}{2} (\exp(-\gamma_0 N) + \exp(-\gamma_1 N)).
\end{aligned}$$

Where the second inequality follows by the separation lemma and Corollary 1. Since $(1+x)^m - 1 \sim mx$ for small x , the first term behaves as $N(N-1)C \exp(-\gamma_3 N)$, which converges to 0. The second term is a polynomial multiplied by an exponentially decaying factor, which also clearly converges to 0. Thus $\mathbb{E}[\sum_K \mathbf{m}_{K,N}] \xrightarrow{N \rightarrow \infty} 0$ as claimed.

Case 2: $\beta_0 < 0$

For this case we divide the proof into three steps.

Step 1: the period-1 graph is empty with overwhelming probability.

Let

$$\mathcal{E}_N := \{D_{ij1} = 0 \ \forall i < j\}$$

denote the event that the period-1 graph is empty.

Fix a dyad ij . Since $\Lambda(x) \leq e^x$ for all $x \in \mathbb{R}$, we have

$$\Pr(D_{ij1} = 1 \mid \mathbf{D}_0, \mathbf{A}) = \Lambda(\alpha_0 D_{ij0} + \beta_0 R_{ij0} + A_{ij}) \leq \exp(\alpha_+ + \beta_0 R_{ij0} + \bar{A}),$$

where $\alpha_+ := \max\{\alpha_0, 0\}$. Taking expectations and using $\beta_0 < 0$,

$$\Pr(D_{ij1} = 1) \leq e^{\alpha_+ + \bar{A}} \mathbb{E}[e^{\beta_0 R_{ij0}}].$$

Under $\mathbf{D}_0 \sim G(N, p)$,

$$R_{ij,0} \sim \text{Bin}(N-2, p^2),$$

hence

$$\mathbb{E}[e^{\beta_0 R_{ij,0}}] = (1 - p^2 + p^2 e^{\beta_0})^{N-2} \leq \exp(-(1 - e^{\beta_0})(N-2)p^2).$$

Therefore there exist constants $C_1, C_2 > 0$ such that, for all N large enough,

$$\Pr(D_{ij1} = 1) \leq C_1 e^{-C_2 N p^2}.$$

By a union bound,

$$\Pr(\mathcal{E}_N^c) \leq \sum_{i < j} \Pr(D_{ij1} = 1) \leq C_1 N^2 e^{-C_2 N p^2}.$$

Since $N p^2 / \log N \rightarrow \infty$, it follows that

$$\Pr(\mathcal{E}_N^c) = o(N^{-m}) \quad \text{for every fixed } m > 0.$$

Step 2: on $\mathcal{E}_N$, a given candidate star is exponentially unlikely to be identifying.

Fix $K \in \{2, \dots, N-1\}$ and consider a candidate K -star system

$$\mathcal{S} = \{i_1 i_2, \dots, i_1 i_K\}$$

with center i_1 and leaves $i_2, \dots, i_K$.

On $\mathcal{E}_N$, we have $D_{uv1} = 0$ for every dyad uv , and hence also

$$R_{uv1} = 0 \quad \text{for every } u \neq v.$$

Therefore, conditional on $(\mathcal{E}_N, \mathbf{A})$, the period-2 link variables $\{D_{uv2}\}_{u < v}$ are independent across dyads and satisfy

$$\Pr(D_{uv2} = 1 \mid \mathcal{E}_N, \mathbf{A}) = \Lambda(A_{uv}), \quad \Pr(D_{uv2} = 0 \mid \mathcal{E}_N, \mathbf{A}) = 1 - \Lambda(A_{uv}) = \Lambda(-A_{uv}).$$

Since $|A_{uv}| \leq \bar{A}$, both probabilities are bounded above by

$$c_* := \Lambda(\bar{A}) < 1.$$

Now suppose that $\mathcal{S}$ is identifying. Because $\mathbf{D}_1 = 0$, Definitions 4 and 5 imply the following period-2 requirements:

1. each of the $K-1$ spoke dyads in $\mathcal{S}$ must satisfy

$$D_{i_1 i_m 2} = 1, \quad m = 2, \dots, K;$$

2. each of the $\binom{K-1}{2}$ leaf-leaf dyads must satisfy

$$D_{i_m i_l 2} = 0, \quad 2 \leq m < l \leq K;$$

3. each dyad joining one of the K star vertices to one of the $N - K$ outside vertices must satisfy

$$D_{uv2} = 0.$$

There are exactly $K(N - K)$ such dyads.

Hence

$$\Pr(\mathcal{S} \text{ is identifying} \mid \mathcal{E}_N, \mathbf{A}) \leq c_*^{L_{N,K}}.$$

with

$$L_{N,K} := (K - 1) + \binom{K - 1}{2} + K(N - K) = K(N - K) + \binom{K}{2} = K \left(N - \frac{K + 1}{2} \right).$$

Step 3: summing over all sizes and all systems.

Define:

$$Y_{ij} := 1\{\text{dyad } ij \text{ is a spoke dyad of some identifying star system}\}.$$

By Dyad Separation (see Corollary 1), any dyad belongs to at most one identifying star system, and every identifying star system contains at least one spoke dyad. Therefore

$$\sum_K \mathbf{m}_{K,N} \leq \sum_{i < j} Y_{ij}.$$

By symmetry,

$$\mathbb{E} \left[\sum_K \mathbf{m}_{K,N} 1(\mathcal{E}_N) \right] \leq \binom{N}{2} \Pr(Y_{12} = 1, \mathcal{E}_N) = \binom{N}{2} \mathbb{E}[1(\mathcal{E}_N) \Pr(Y_{12} = 1 \mid \mathcal{E}_N, \mathbf{A})].$$

Fix K . The number of candidate K -stars for which dyad 12 is a spoke is

$$2 \binom{N - 2}{K - 2},$$

since either node 1 is the center and node 2 is one of the leaves, or vice versa, and then the remaining $K - 2$ leaves are chosen from the other $N - 2$ nodes. Hence, by a union bound

and the estimate from Step 2,

$$\Pr(Y_{12} = 1 \mid \mathcal{E}_N, A) \leq \sum_{K=2}^{N-1} 2 \binom{N-2}{K-2} c_*^{L_{N,K}}.$$

We now bound this sum.

$$\mathbb{E} \left[\sum_K \mathbf{m}_{K,N} \mid \mathcal{E}_N, A \right] \leq N \sum_{K=2}^{N-1} \binom{N-1}{K-1} c_*^{L_{N,K}}$$

For $2 \leq K \leq N/2$, we have

$$L_{N,K} = K \left(N - \frac{K+1}{2} \right) \geq \frac{KN}{2},$$

so

$$\binom{N-1}{K-1} c_*^{L_{N,K}} \leq N^{K-1} c_*^{KN/2} = c_*^{N/2} (N c_*^{N/2})^{K-1}.$$

Since $c_* < 1$, we have $N c_*^{N/2} \rightarrow 0$, and therefore, for all large N ,

$$\sum_{K=2}^{\lfloor N/2 \rfloor} N \binom{N-1}{K-1} c_*^{L_{N,K}} \leq \sum_{K=2}^{\lfloor N/2 \rfloor} N \binom{N-1}{K-1} (N c_*^{N/2})^K \leq 2N^3 c_*^N.$$

where the last inequality follows by upper bounding the sum from $K = 2$ to $\lfloor N/2 \rfloor$ by the sum from $K = 2$ to $N - 1$ then using the binomial identity.

For $K > N/2$, the function $K \mapsto L_{N,K}$ is bounded below by a quadratic term; in particular, for all large N ,

$$L_{N,K} \geq \frac{N^2}{4}.$$

Hence

$$\sum_{K=\lfloor N/2 \rfloor + 1}^{N-1} N \binom{N-1}{K-1} c_*^{L_{N,K}} \leq N 2^N c_*^{N^2/4}.$$

Combining the two ranges,

$$\mathbb{E} \left[\sum_K \mathbf{m}_{K,N} \mid A, \mathcal{E}_N \right] \leq N 2^N c_*^{N^2/4} + 2N^3 c_*^N$$

Note that

$$\begin{aligned}\mathbb{E} \left[\sum_K \mathbf{m}_{K,N} | \mathbf{A} \right] &= \mathbb{E} \left[\sum_K \mathbf{m}_{K,N} | \mathbf{A}, \mathcal{E}_N \right] \Pr(\mathcal{E}_N) + \mathbb{E} \left[\sum_K \mathbf{m}_{K,N} 1(\mathcal{E}_N^c) | \mathbf{A} \right] \\ &\leq N 2^N c_*^{N^2/4} + 2N^3 c_*^N + N \times \Pr(\mathcal{E}_N^c | \mathbf{A})\end{aligned}$$

where the inequality follows from the observation that $\sum_K \mathbf{m}_{K,N} \leq N$ almost surely, by the separation lemma. Taking expectations, and recalling that $c_* < 1$, we get:

$$\mathbb{E} \left[\sum_K \mathbf{m}_{K,N} \right] \leq N 2^N c_*^{N^2/4} + 2N^3 c_*^N + \Pr(\mathcal{E}_N^c) \rightarrow 0,$$

as claimed.

B.7 Proof of Proposition 3

For any dyad ij ,

$$R_{ijt} \leq \#\{k \neq i, j : \|z_k - z_i\| \leq r, \|z_k - z_j\| \leq r\}.$$

The set on the right is contained in the intersection of two radius- r balls, hence in a region of area at most πr^2 . Since the square has area N , that dominating count is $\text{Binomial}(N-2, p)$ with $p \leq \pi r^2/N$, so it is tight, indeed $O_p(1)$. Thus, in sparse geometric graphs, the raw common-neighbor count is already locally bounded in probability.

We begin with the upper bound. If $\mathcal{S}_i$ is identifying, then every spoke $i_1 i_m$ must be feasible, because Definition 5 requires

$$D_{i_1 i_m 1} \neq D_{i_1 i_m 2},$$

and that is impossible when $A_{i_1 i_m} = -\infty$. Hence

$$\Pr(Z_{\mathbf{i},K} = 1) \leq \Pr(\|z_{i_1} - z_{i_m}\| \leq r, m = 2, \dots, K).$$

Let $B(x, r)$ be the closed ball in $\mathbb{R}^2$ centered at x with radius r . Conditioning on $z_{i_1} = x$,

$$\Pr(\|z_{i_1} - z_{i_m}\| \leq r \ \forall m \mid z_{i_1} = x) = \left(\frac{|B(x, r) \cap [0, \sqrt{N}]^2|}{N} \right)^{K-1} \leq \left(\frac{\pi r^2}{N} \right)^{K-1}.$$

Therefore

$$\Pr(Z_{\mathbf{i},K} = 1) \leq \left(\frac{\pi r^2}{N} \right)^{K-1}.$$

We next turn to the lower bound. Let

$$M := \binom{K}{2},$$

the number of internal dyads among the K star nodes. Consider the event E_i defined by:

1. the center is well inside the square:

$$z_{i_1} \in [3r/2, \sqrt{N} - 3r/2]^2;$$

2. each leaf lies within distance $r/2$ of the center:

$$z_{i_m} \in B(z_{i_1}, r/2), \quad m = 2, \dots, K;$$

3. no other agent lies in $B(z_{i_1}, 3r/2)$;

4. among the M feasible internal dyads:

- at $t = 0$, all are absent;
- at $t = 1$, all possible edges are absent;
- at $t = 2$, exactly the $K-1$ “spoke-edges” are present and the remaining $M-(K-1)$ “leaf-leaf edges” are absent.

On E_i , the candidate K -star is identifying:

- all spokes switch $1 \rightarrow 2$, so Definition 5(ii) holds;
- because all leaves are within $r/2$ of the center, any outside node that could connect to any star node would have to lie within $3r/2$ of the center, but event 3 rules that out;
- hence the star has no outside feasible neighbors at either $t = 1$ or $t = 2$;
- the period-1 neighborhood is empty, so Definition 4(ii) is vacuous;
- all non-spoke internal dyads are stable (absent) in $t = 1$ and $t = 2$, so Definition 4(iii) holds.

Thus

$$E_i \subseteq \{Z_{i,K} = 1\}.$$

We now compute the probability of $E_{\mathbf{i}}$. First, for large N ,

$$\Pr\left(z_{i_1} \in [3r/2, \sqrt{N} - 3r/2]^2\right) = \frac{(\sqrt{N} - 3r)^2}{N} \geq \frac{1}{2}.$$

Second, conditional on that interior event and on z_{i_1} ,

$$\Pr\left(z_{i_m} \in B(z_{i_1}, r/2), m = 2, \dots, K \mid z_{i_1}\right) = \left(\frac{\pi r^2}{4N}\right)^{K-1}.$$

Third, again conditional on the center being interior and the leaves lying in $B(z_{i_1}, r/2)$,

$$\Pr\left(\text{no other agent in } B(z_{i_1}, 3r/2)\right) = \left(1 - \frac{9\pi r^2}{4N}\right)^{N-K},$$

which is bounded below by a positive constant for all large N .

So the geometry part contributes

$$\Pr(\text{items 1-3}) \geq c_{1,K} N^{-(K-1)}$$

for some $c_{1,K} > 0$.

Finally, conditional on items 1-3, all feasible dyads are internal to the K nodes. On the event $E_{\mathbf{i}}$, all internal dyads are zero at $t = 0$ and $t = 1$, so every relevant common-neighbor count is $R_{ij0} = R_{ij1} = 0$.

The probability of item 4, conditional on items 1-3, is

$$(1 - \Lambda(a))^M \cdot (1 - \Lambda(a))^M \cdot \Lambda(a)^{K-1} (1 - \Lambda(a))^{M-(K-1)},$$

which is a strictly positive constant depending only on K and a . (Recall that $A_{ij} = a$ when nodes i and j are distance r or less apart.) Therefore,

$$\Pr(Z_{\mathbf{i},K} = 1) \geq \Pr(E_{\mathbf{i}}) \geq c_{1,K} (1 - \Lambda(a))^M \cdot (1 - \Lambda(a))^M \cdot \Lambda(a)^{K-1} (1 - \Lambda(a))^{M-(K-1)} N^{-(K-1)}.$$

This proves

$$\Pr(Z_{\mathbf{i},K} = 1) \asymp N^{-(K-1)}.$$

To conclude, note that the number of candidate K -stars is

$$N \binom{N-1}{K-1},$$

because the center is distinguished and the $K - 1$ leaves are chosen from the remaining $N - 1$ agents. Therefore

$$\mathbb{E}[\mathbf{m}_{K,N}] = N \binom{N-1}{K-1} \Pr(Z_{i,K} = 1) \asymp N \cdot N^{K-1} \cdot N^{-(K-1)} \asymp N.$$

Hence the expected number of identifying K -systems is of order N . Summing over $K = 2, \dots, \bar{K}$ yields

$$\mathbb{E} \left[\sum_{K=2}^{\bar{K}} \mathbf{m}_{K,N} \right] \asymp N$$

for every fixed $\bar{K}$. □