Homophily and transitivity in dynamic network formation

Bryan S. Graham*

April 2016

Abstract

In social and economic networks linked agents often share additional links 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 links in common. Second, agents may assortatively match on unobserved attributes, a process called homophily. I study parameter identifiability in a simple model of dynamic network formation with both effects. Agents form, maintain, and sever links over time in order to maximize utility. The return to linking may be higher if agents share friends in common. A pair-specific utility component allows for arbitrary homophily on time-invariant agent attributes. I derive conditions under which it is possible to detect the presence of a taste for transitivity in the presence of assortative matching on unobservables. I leave the joint distribution of the initial network and the pair-specific utility component, a very high dimensional object, unrestricted. The analysis is of the 'fixed effects' type. The identification result is constructive, suggesting an analog estimator, whose single large network properties I characterize.

JEL Codes: C31, C33, C35

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

*Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3888, e-mail: bgraham@econ.berkeley.edu, web: http : //bryangraham.github.io/econometrics/.
This work builds on 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). I am grateful to Guido, and then JBES editors Kei Hirano and Jonathan Wright, for providing the initial inspiration for this project. I am thankful to Michael Leung for introducing me to the literature on random geometric graphs. I am also grateful to Peter Bickel, Paul Goldsmith-Pinkham, Jim Heckman, Bo Honore, Noureddine El Karoui 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) and the USC-INET Conference on Networks (November 20-21st, 2015) for valuable comments. Financial support from NSF Grant SES #1357499 is gratefully acknowledged. All the usual disclaimers apply.

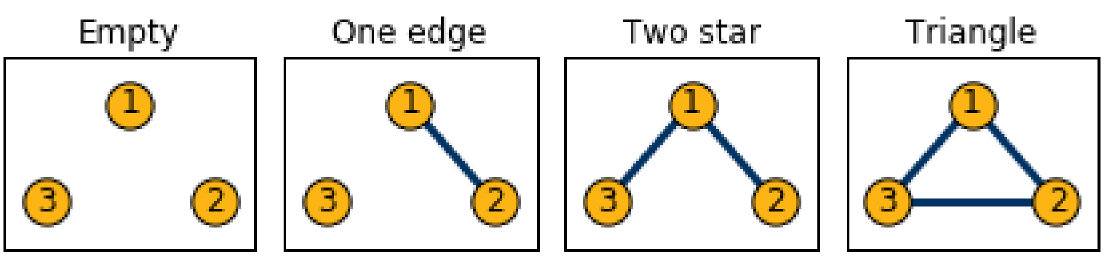
Let i, j and k index three independent random draws from some network of agents (i.e., a population of potentially connected sampling units). Together these agents constitute a *triad*. Let $D_{ij} = 1$ if agents i and j are connected and zero otherwise. Links are undirected, such that $D_{ij} = D_{ji}$ for all pairs ij, and self-ties are ruled out, so that $D_{ii} = 0$ for all i.

A triad can be wired in one of four ways (see Figure 1). Consider the probability of the observing the wiring where all three pairs, ij, ik and jk, are connected conditional on at least two of the pairs being connected:

$$\rho_{\mathrm{CC}} = \Pr\left(D_{ij} = 1 \middle| D_{ik} = 1,\, D_{jk} = 1\right) \tag{1}$$

The sample analog of (1) is called the transitivity index or global clustering coefficient in the networks literature (Graham, 2015). In real world social and economic networks ρ_{CC} is generally higher than $\rho_{\mathrm{D}} = \Pr(D_{ij} = 1)$, the unconditional frequency at which agents link (Jackson, 2008). It is often substantially higher. Links are clustered. Networks exhibit transitivity: agents are more likely to link if they share links in common ("the friend of my friend is also my friend").

Figure 1: Triad configurations in undirected networks



High levels of transitivity are found in friendship networks, industrial supply-chains, international trade flows, and alliances across firms and nations (Gulati and Gargiulo, 1999; Choi and Wu, 2009; Maoz, 2012; Matous and Todo, 2016; Davis et al., 1971; Kossinets and Watts, 2009).[1] In his classic paper on weak ties, Granovetter (1973), in honor of their supposed infrequent occurrence, even calls the intransitive two-star a "forbidden triad" (see Figure 1).[2]

Transitivity in links may arise for two distinct reasons. First agents may have a *structural taste* for transitive ties. The returns to link formation between any two agents may be increasing in the number of neighbors they share in common. Coleman (1990) argues

[1] Jackson (2008) gives many other examples and references.

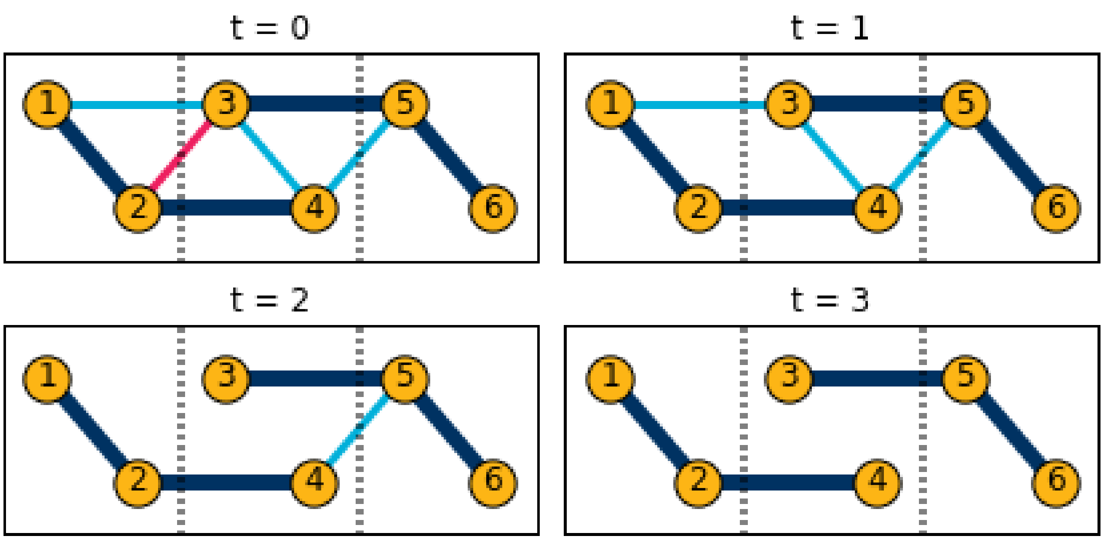
[2] Davis et al. (1971), in another classic reference, claim that transitivity of social ties is the "central proposition in structural sociometry" (p. 309).

that closed networks facilitate the formation of social and human capital.[3] This might provide an impetus for agents to form transitive relationships. 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 in social ties may reflect assortative matching on an unobserved attribute. That is, the returns to linking 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).

Distinguishing between a structural taste for transitivity and homophily as drivers of clustering in networks is of considerable scientific interest and policy relevance. In the presence of a structural taste for transitivity outside interventions can have long run effects on network structure. If clustering primarily reflects assortative matching on unobserved agent attributes, interventions are less likely to lead to durable changes in network structure.

Figure 2: The effects of link deletion when agents prefer transitive ties



Notes: The vertical dashed lines partition nodes into three distinct communities. Agents may only form ties with agents in their own and adjacent communities. The thick "Berkeley Blue" lines correspond to edges that form regardless of network structure. The thinner "Lawrence" colored lines correspond to edges that form only when the two agents share another link in common. The thin "Rose Garden" colored edge its targeted for deletion by the policy-maker.

[3] For example both parents and teachers interact with a student, but if parents and teachers also consult with each other about the student, forming a triangle, the student's rate of learning may increase.

Figure 2 provides one illustration of these claims. Agents are divided into three communities according to the vertical dashed lines in the figure. Links can only arise between agents in the same or adjacent communities (i.e., there exists locational homophily). There are two types of links, "strong ties", which are sustainable regardless of network structure (thick darker edges) and "weak ties", which require the support of a link in common (thin lighter edges). The initial network structure is given in the upper-left-hand $t = 0$ graph. This network consists of a single "giant component".

Now consider a policy-maker who deletes the link between agents 2 and 3 (the thin "Rose Garden" colored edge). The immediate effect of this deletion is depicted in the $t = 1$ graph in the upper-right-hand corner of the figure. Once this link is deleted pairs 1 and 3 and 3 and 4 no longer share a friend in common. Since these edges require a common link for sustenance, they dissolve, leading to the $t = 2$ graph depicted in the lower-left-hand corner of the figure. However, since now agents 4 and 5 no longer share a link in common, their tie also dissolves leading to the $t = 3$ network configuration depicted in the lower-right-hand corner of the figure. The architecture of the final graph is very different from the initial one, consisting of two separate components.

The presence of a structural taste for transitivity creates interdependencies in link formation, whereby the surplus agents i and j get from linking may vary with the presence of absence of edges elsewhere in the network. The presence of such interdependencies means that a local manipulation may have global consequences for network structure. Removing or adding a single link may initiate a cascade of edge removals and/or additions.

When interdependencies are absent, the effect of deleting a link would be entirely local. In the example, the final state of the network in such a case would be given by the $t = 1$ graph in the upper-right-hand corner of Figure 2. This graph does not appreciably differ from its pre-policy $t = 0$ version.

In this paper I introduce an empirical model of dynamic network formation and formally study its identifiability. I am particularly interested in circumstances under which it is possible to attribute clustering in ties, a central feature of real world networks, to a *structural taste* for transitivity versus assortative matching on an unobserved attribute (i.e., *homophily*).

The motivation for my question is twofold. First, networks are ubiquitous and their structure is evidently important for a variety of social and economic outcomes (Jackson et al., 2016). Second, as argued using the example in Figure 2, the manipulability of network structure hinges on the degree to which current (local) network structure influences future (global) network structure. While the homophily versus transitivity identification problem has been informally articulated in the networks literature (e.g., Gulati and Gargiulo, 1999; Goodreau et al., 2009) it has not been systematically analyzed.

There is a useful analogy between my research problem, both in terms of scientific and policy 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 my problem makes it more challenging, nevertheless I will utilize intuitions and proof strategies from this earlier research as well as from Manski's semiparametric analysis of discrete choice (Manski, 1975, 1985), especially as extended to static (Manski, 1987) and dynamic (Honoré and Kyriazidou, 2000) panel data models .

Although network analysis has a rich empirical history in sociology and other disciplines, there are, as yet, few empirical models of strategic network formation.[4] In strategic models agents form, maintain, or sever links with each other in order to maximize utility (Jackson and Watts, 2002; Jackson and Wolinsky, 1996; Watts, 2001). This approach to econometrically modeling network formation follows that pioneered by McFadden (1974) for single agent discrete choice problems.

The multi-agent nature of networks complicates structural model-building. At the risk of some simplification and omission, recent work in econometrics has taken one of three different, albeit complementary, approaches.[5]

The first approach ignores interdependencies in link formation, focusing instead, on the introduction of rich forms of correlated heterogeneity into dyadic models of link formation (Dzemski, 2014; Graham, 2014; Krivitsky et al., 2009; van Duijn et al., 2004). This corresponds to the network analog of studying identification and estimation in single agent *static* binary choice models (Chamberlain, 1980; Manski, 1987).

The second approach ignores heterogeneity, focusing instead, on interdependencies in link formation. This approach borrows from earlier work on the econometrics of games, but introduces new insights to handle the combinatoric complexity of the many player network formation game (Christakis et al., 2010; Sheng, 2014; de Paula et al., 2015; Mele, 2015; Menzel, 2015).

A third approach, and the one also adopted here, explores the identifying value of multiple observations of network structure (Snijders, 2011; Goldsmith-Pinkham and Imbens, 2013; Graham, 2013).

I assume that the econometrician observes a single network over multiple periods. In each

[4] Graham (2015), de Paula (2016) and Chandrasekhar (2016) provide recent reviews of the econometric literature on network formation. Goldenberg et al. (2010) is a recommended review of the statistics and machine learning literature.

[5] For example Chandrasekhar and Jackson (2015) does not fit easily into this typology.

period agents form, maintain, or sever links taking the structure of the network in the preceding period as fixed. The systematic component of link utility varies with current network structure. Agents are not forward-looking, rather they treat the current structure of the network as fixed when deciding whether to maintain, sever or form a link. This is analogous to the simple best-reply dynamic used to explore issues of existence and stability in theoretical work (Jackson and Watts, 2002; Jackson and Wolinsky, 1996; Watts, 2001), and also parallels the empirical approach taken in Christakis et al. (2010), Snijders (2011) and Goldsmith-Pinkham and Imbens (2013).

My treatment of unobserved heterogeneity in this setting is innovative. I divide the unobserved component of link utility into a time-invariant pair-specific component and a time-varying 'idiosyncratic' component. The latter is independently and identically distributed over time within pairs of agents and independently, although not necessarily identically, distributed across pairs. The distribution of the pair-specific component is left unrestricted. In a network with N agents there will be a total of $n = \frac{1}{2}N(N-1)$ pair-specific heterogeneity terms. This is a high-dimensional latent variable. My set-up leaves its joint distribution unrestricted. Furthermore I leave its relationship with the initial network condition unrestricted. Put differently, in contrast to all prior work of which I am aware, my identification analysis is of the 'fixed effects', as opposed to 'random effects', type (cf., Chamberlain, 1985; Honoré and Tamer, 2006).

The upshot of these innovations is the first empirical model of strategic network formation that incorporates both interdependencies in link formation *and* rich forms of correlated heterogeneity. This setting, in turn, allows for the first precise articulation of, and test for, the homophily versus transitivity hypothesis.

The next section introduces a semiparametric model of dynamic network formation. Section 2 considers identifiability of the model's parameters. Section 3 presents estimation results, including large sample theory. Section 4 summarizes the results of some Monte Carlo experiments. Section 5 discusses possible extensions. Proofs and some supplemental results are collected in Appendices A, and B. Replication code for the Monte Carlo experiments is available in the supplementary materials.

Notation

In what follows various sets of agents and dyads will play a prominent role. In defining these sets it is useful to recall that "$\vee$" denotes the non-exclusive "or", "$\underline{\vee}$" denotes the exclusive "or", "$\exists!$" denotes "there exists exactly one".

Random variables are denoted 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. I use ι_N to denote a $N \times 1$ vector of ones, I_N the $N \times N$ identity matrix, and $A \circ B$ denotes the Hadamard (i.e., entry-wise) product of the conformable matrices A and B. A "0" subscript on a parameter denotes its population value and may be omitted when doing so causes no confusion.

1 A dynamic model of network formation

Consider a group of N, potentially connected individuals indexed by i. Individuals may be equivalently referred to as agents, players, nodes or vertices depending on the context. We observe all ties across members in each of $t = 0, \ldots, T$ periods. Recall that $D_{ijt} = 1$ if agents i and j are linked in period t and zero otherwise. Links may be equivalently referred to as friendships, connections, edges or arcs depending on the context.

Let $R_{ijt} = \sum_{k=1}^{N} D_{ikt} D_{jkt}$ equal the number of links i and j have in common in period t. Individuals i and j form a link in period $t = 1, \ldots, T$ according to

$$D_{ijt} = \mathbf{1}\left(\beta_0 D_{ijt-1} + \gamma_0 R_{ijt-1} + A_{ij} - U_{ijt} \geq 0\right), \tag{2}$$

where $\mathbf{1}(\bullet)$ is the indicator function, A_{ij} is a pair-specific, time invariant, unobserved heterogeneity term, and U_{ijt} a time-varying shock. Since links are undirected $A_{ij} = A_{ji}$ and similarly for U_{ijt}. We can think of the term inside the indicator function as the net surplus associated with an ij link. Rule (2) implicitly conceptualizes utility as being transferrable across agents (cf., Graham, 2011). Goldsmith-Pinkham and Imbens (2013) specify a related data generating process that assumes non-transferable utility.

Rule (2) implies a link between i and j is more likely if (i) they were linked in the previous period ($D_{ijt-1} = 1$), (ii) they shared many links in common last period (R_{ijt-1} is large) or (iii) there are unobserved pair attributes which generate large surplus from linking (i.e., A_{ij} is large).[6]

Let $\mathbf{D}_t$ be the $N \times N$ matrix with D_{ijt} as its ij^{th} element. Note that $\mathbf{D}_t$ is symmetric and has a diagonal consisting of zeros. This matrix is called the period t network adjacency matrix. Let $\mathbf{A}$ be the $N \times N$ matrix of pair heterogeneity terms (i.e., the matrix with ij^{th} element A_{ij}). Rule (2) only applies to periods 1 to T. I leave the joint distribution of $(\mathbf{D}_0, \mathbf{A})$, the

[6] Here, as well as in what follows, I assume, for expository purposes, that both β and γ are positive.

initial condition of the network sequence, unrestricted:

$$\left(\mathbf{D}_0, \mathbf{A}\right) \sim \Pi_0. \tag{3}$$

The density function evaluated at $\mathbf{D_0} = \mathbf{d}_0$, $\mathbf{A} = \mathbf{a}$ is denoted by $\pi_0\left(\mathbf{d}_0, \mathbf{a}\right)$.

To close the model we assume that (i) U_{ijt} is independently and identically distributed over time and (ii) independently, although not necessarily identically, distributed across pairs[7]:

$$F\left(U_{121}, \ldots, U_{12T}, \ldots, U_{N-1N1}, \ldots, U_{N-1NT}\right) = \prod_{i<j} \prod_{t=1}^{T} F_{U,ij}. \tag{4}$$

Throughout I assume that $F_U\left(u\right)$ is strictly increasing on $\mathbb{R}^1$. Here, and in what follows, I use the notation $\prod_{i<j}$ to indicate $\prod_{i=1}^{N}\prod_{j=i+1}^{N}$ and similarly $\sum_{i<j}$ to indicate $\sum_{i=1}^{N}\sum_{j=i+1}^{N}$. Rule (2) parsimoniously captures three forces hypothesized by researchers as important for link formation (cf., Snijders, 2011, 2013). First, there is *state dependence* in links; all things equal the returns to linking for i and j are higher in the current period if they were also connected in the prior period. Second, there are returns to 'triadic closure'; my utility is higher if the "friends of my friends are also my friends". 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 individual-specific characteristics and $g\left(\xi_i, \xi_j\right)$ a measure of the distance between i and j in ξ (i.e., $g\left(\bullet, \bullet\right)$ is a distance function). If $A_{ij} = -g\left(\xi_i, \xi_j\right)$, then rule (2) 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

$$A_{ij} = \nu_i + \nu_j - g\left(\xi_i, \xi_j\right), \tag{5}$$

allows for the possibility that certain individuals may generate high friendship surplus. Put differently ν_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 (cf., Krivitsky et al., 2009; Graham, 2014).

Both a *structural* taste for transitivity in relationships, here parameterized by γ, and ho-

[7] In what follows I will generally suppress the potential dependence of F_U on ij.

mophily on unobserved attributes, can generate high levels of clustering in networks.

Likelihood

Equations (2), (3) and (4) specify a dynamic model of network formation. The joint probability density of a specific sequence of network configurations $\mathbf{D}_0^T = (\mathbf{D}_0, \mathbf{D}_1, \ldots, \mathbf{D}_T)$ and realization of $\mathbf{A}$ is

$$\begin{aligned} p\left(\mathbf{d}_0^T, \mathbf{a}, \theta\right) &= \pi\left(\mathbf{d}_0, \mathbf{a}\right) \\ &\times \prod_{i<j} \prod_{t=1}^{T} F\left(\beta d_{ijt-1} + \gamma r_{ijt-1} + a_{ij}\right)^{d_{ijt}} \\ &\times \left[1 - F\left(\beta d_{ijt-1} + \gamma r_{ijt-1} + a_{ij}\right)\right]^{1-d_{ijt}} \end{aligned} \tag{6}$$

Let $\pi_{\mathbf{D}_0|\mathbf{A}}\left(\mathbf{d}_0 | \, \mathbf{a}\right) = \Pr\left(\mathbf{D}_0 = \mathbf{d}_0 | \, \mathbf{A} = \mathbf{a}\right)$ denote the distribution of the initial network given $\mathbf{A} = \mathbf{a}$. Let $\pi_{\mathbf{A}}\left(\mathbf{a}\right)$ denote the marginal density function for $\mathbf{A}$. The integrated likelihood for the observed data is then

$$\begin{aligned} p^{\mathrm{I}}\left(\mathbf{d}_0^T, \theta\right) &= \int \cdots \int \prod_{t=1}^{T} \prod_{i<j} \Big\{ F_U\left(\beta d_{ijt-1} + \gamma r_{ijt-1} + a_{ij}\right)^{d_{ijt}} \\ &\quad \times \left[1 - F_U\left(\beta d_{ijt-1} + \gamma r_{ijt-1} + a_{ij}\right)\right]^{1-d_{ijt}} \Big\} \\ &\quad \times \pi_{\mathbf{D}_0|\mathbf{A}}\left(\mathbf{d}_0 | \, \mathbf{a}\right) \pi_{\mathbf{A}}\left(\mathbf{a}\right) \mathrm{d}a_{21}, \ldots, \mathrm{d}a_{NN-1}. \end{aligned} \tag{7}$$

Even if $F_U\left(\bullet\right)$ and $\pi_{\mathbf{A}}\left(\bullet\right)$ are parametrically specified, using (7) as a basis for estimation and inference is problematic. Three issues arise.

The first is familiar from prior work on dynamic discrete choice analysis: rule (2) provides no guidance on how to specify $\pi_{\mathbf{D}_0|\mathbf{A}}\left(\mathbf{d}_0 | \, \mathbf{a}\right)$ (Heckman, 1981c; Honoré and Tamer, 2006). In analogy I call this the *initial network problem*. Rule (2) 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 $\pi_{\mathbf{D}_0|\mathbf{A}}\left(\mathbf{d}_0 | \, \mathbf{a}\right)$ coincides with the steady state distribution implied by (2) (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 $\pi_{\mathbf{D}_0|\mathbf{A}}\left(\mathbf{d}_0 | \, \mathbf{a}\right)$ were correctly specified, evaluating (7) requires computing a very high-dimensional integral. Since it seems reasonable, at the very minimum, to choose a specification for $\pi_{\mathbf{A}}\left(\mathbf{a}\right)$ which allows A_{ij} and A_{ik} to covary, there is no obvious way to factor (7) to reduce the dimensionality of the required integration. Goldsmith-Pinkham and Imbens (2013) assume that $A_{ij} = \alpha_\xi \left|\xi_i - \xi_j\right|$ 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 $\left(\hat{\beta}_{ML}, \hat{\gamma}_{ML}\right)$; 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 approach to inference developed below, in contrast, is frequentist.

2 Fixed effects identification

With $F_U(\bullet)$ and $\pi(\bullet)$ (semi-) parametrically specified, an approach to estimation and inference based on the integrated likelihood (7), whether frequentist or Bayesian, is a so called *random effects* one. The discussion above indicates that the random effects approach to network analysis involves delicate modeling issues as well as serious computational challenges.

In this section I explore the *fixed effects* identifiability of (β, γ). That is I explore what can be learned about these two parameters when the joint distribution of $(\mathbf{D}_0, \mathbf{A})$ is left completely unrestricted. The fixed effects framework is a natural one in which to begin any formal analysis of identification, since adding restrictions to the model, as the random effects approach does, can only improve identifiability.

Nevertheless one might question the fruitfulness of a fixed effects analysis at the outset. The random variable $(\mathbf{D}_0, \mathbf{A})$ is of very high dimension, consisting of $N(N-1)$ components. This suggests that leaving its joint distribution unspecified may lead to serious identification problems. A key contribution of this paper is to show that a fixed effect approach to dynamic network analysis is feasible and fruitful.

In single agent models fixed effects identification of true state dependence in the presence of unobserved heterogeneity is based on the frequency of observing certain sequences of choices relative to other sequences (e.g., Cox, 1958; Heckman, 1978; Chamberlain, 1985; Honoré and Kyriazidou, 2000). For example, in the absence of state dependence the binary sequences 0101 and 0011 are equally likely. In the presence of state dependence, the relative frequency of the latter sequence will be greater (under certain assumptions).

The approach to identification developed below is similar, being based on the relative frequency of certain ij friendship histories. However the analysis is more complicated than

[8] van Duijn et al. (2004) introduce MCMC methods that also might be adapted in order to evaluate (7) in special cases.

in the single agent case. This is because the effect of the pair ij forming or terminating a link in period t cascades through the period $t+1$ portion of the likelihood. Such an effect was implicit in the discussion accompanying Figure 2 above. For example if ik are linked in period t, then the addition of an ij link increases the probability of a jk link in period $t+1$. Local changes in the network can have widespread effects on the structure of the network in subsequent periods.

The high level of interdependence across different pairs of agents' link decisions makes the mapping from the frequency of relative friendship histories to the model parameters less direct than in the single agent case. My approach is to consider pairs of agents that are embedded in a stable neighborhood (defined below). The relative frequencies of different friendship histories, when pairs are themselves embedded in different types of stable neighborhoods, provides information about the model parameters. Intuitively my approach involves making comparisons 'holding other features of the network fixed'. This is not straightforward to do. The likelihood functions associated with two network histories, identical in all respects except that the ij friendship history in one is a permutation of that in the other, may be very different due to the interdependent nature of linking in the model.

Stable neighborhoods

I consider the $T = 3$ case. This corresponds to four network observations, with the initial network receiving a '0' subscript (this is the minimal number of observations needed for a positive result). As will become apparent shortly, the extension to the general case is straightforward (cf., Charlier et al., 1995)

The period t *neighbors* of dyad ij are given by the index set

$$n_t(ij) = \{k \mid d_{ik} = 1 \vee d_{jk} = 1\} \setminus \{i, j\}.$$

Dyad ij's neighborhood consists of all agents to which it is directly connected; excluding each other.

Definition 1. (Neighborhood Stability) ij is embedded in a *stable neighborhood* if
(i) $n_1(ij) = n_2(ij) = n_3(ij)$,
(ii) for all $k \in n_1(ij)$ and $l = 1, \ldots, N$; $l \neq i, j$ we have $d_{kl1} = d_{kl2}$.

A stable neighborhood has two features. First, with the exception of possible link formation and dissolution between themselves, the set of links maintained by agents i and j is constant across periods 1, 2 and 3. Constancy in i and j's links across periods 0 and 1 is not required.

To the contrary, some inter-period variation in these links is required for identification. Second, the links maintained by neighbors of players i and j do not change between periods 1 and 2.

Neighborhood stability imposes some time constancy of links up to two degrees away from the reference dyad. Examples of dyads embedded in stable neighborhoods are given in Figures 3 to 5 below.

Definition 2. (STABLE DYAD) Dyad ij is stable if (i) it is embedded in a *stable neighborhood* and (ii) i and j revise their link status between periods 1 and 2 such that $d_{ij1} \neq d_{ij2}$.

A basic implication of Definition 2 is that two stable dyads are separated in a certain sense, a fact that will prove helpful below.

Lemma 1. *(*SEPARATION*) Let ij and kl denote two stable dyads, then the distance between i and k (or equivalently i and l or j and k) is at least two degrees in periods 1 and 2.*

Proof. Suppose i and k are distance one apart, then the second condition for neighborhood stability in Definition 1 is violated for both ij and kl. □

Appendix B outlines how to construct an indicator for dyad stability. A Python 2.7 function for this purpose is included in the supplemental materials. This program makes extensive use of the broadcasting rules embedded in the Numpy module. It can efficiently find all stable dyads in networks consisting of tens of thousands of agents on a desktop computer.

In what follows the binary variable Z_{ij} will equal one if ij is a stable dyad and zero otherwise. It is also helpful to have an index notation for dyads. Recall that $i = 1, 2, \ldots$ indexes the N agents in the network. Let the boldface indices $\mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots$ index the $n = \binom{N}{2} = \frac{1}{2}N(N-1)$ *dyads* among them (in arbitrary order). In an abuse of notation, also let $\mathbf{i}$ denote the set $\{i_1, i_2\}$, where i_1 and i_2 are the indices for the two agents which comprise dyad $\mathbf{i}$. Using this notation we have, for example, $Z_{\mathbf{i}} = Z_{i_1 i_2}$.

Finally the arguments presented below involve various partitions of the N agents and n dyads in the network into different sets. Let $\mathcal{N}_s = \{i | \exists! \, j \, , Z_{ij} = 1\}$ denote the set of all nodes that are part of a stable dyad. Let $\mathcal{N}_s^c = \mathcal{N} \setminus \mathcal{N}_s$ denote the absolute complement of $\mathcal{N}_s$ in $\mathcal{N}$; or the set of all nodes which are not part of a stable dyad. Let $\mathcal{D}_s = \{\mathbf{i} \,|\, Z_{i_1 i_2} = 1\}$ denote the set of all stable dyads, $\mathcal{D}_{ns} = \{\mathbf{i} \,|\, i_1 \in \mathcal{N}_s \veebar i_2 \in \mathcal{N}_s\}$ the set of all dyads that are not stable, but include a node who is part of a stable dyad, and $\mathcal{D}_{od} = \mathcal{D} \setminus (\mathcal{D}_s \cup \mathcal{D}_{ns})$ the set of all dyads where both nodes are not part of $\mathcal{N}_s$ (i.e., "all other dyads").

Finally, let $\mathbb{D}$ denote the set of all valid binary undirected adjacency matrices.

Main identification result

Consider the set of network sequences

$$\begin{aligned}\mathbb{V}^s &= \Big\{ \mathbf{v}_0^3 = (\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \Big| \; \mathbf{v}_t \in \mathbb{D} \, \text{for} \, t = 0, \ldots, 3, \\ &\quad \mathbf{v}_0 = \mathbf{d}_0, \, \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{d}_1 + \mathbf{d}_2, \, \mathbf{v}_3 = \mathbf{d}_3, \\ &\quad v_{ij1} = d_{ij1} \,\&\, v_{ij2} = d_{ij2} \, \text{if} \, z_{ij} = 0, \, \text{for} \, i, j = 1, \ldots, N \Big\} . \end{aligned} \tag{8}$$

The set $\mathbb{V}^s$ contains all network sequences constructed by permutating the period 1 and 2 link decisions of the $\mathbf{m}_N \overset{def}{\equiv} |\mathcal{D}_s|$ dyads embedded in stable neighborhoods. All other link decisions are held fixed at their observed values. The set $\mathbb{V}^s$ therefore contains $2^{|\mathcal{D}_s|} = 2^{\mathbf{m}_N}$ elements. To restate, it consists of all network sequences generated by permutating the observed of period 1 and 2 link decisions of dyads that (i) revise their linking behavior across periods 1 and 2 and (ii) are embedded in stable neighborhoods. All other link decisions coincide with the observed ones.

The main result is a conditional likelihood expression.

Theorem 1. *(*CONDITIONAL LIKELIHOOD*) Under the data generating process specified in Section 1 the conditional likelihood of the event $D_0^3 = \mathbf{d}_0^3$ given that $\mathbf{d}_0^3 \in \mathbb{V}^s$,*

$$l^c \left(\mathbf{d}_0^3, \mathbf{a}, \theta \right) = \frac{p \left(\mathbf{d}_0^3, \mathbf{a}, \theta \right)}{\sum_{\mathbf{v} \in \mathbb{V}^s} p \left(\mathbf{v}_0^3, \mathbf{a}, \theta \right)}, \tag{9}$$

equals

$$l^c \left(\mathbf{d}_0^3, \mathbf{a}, \theta \right) = \prod_{\mathbf{i} \in \mathcal{D}_s} \left[\frac{1}{1 + b^{01}_{i_1 i_2} \left(q_{ij}, a_{ij}, \theta \right)} \right]^{\mathbf{1}\left(s_{i_1 i_2} = 1\right)} \left[\frac{1}{1 + b^{10}_{i_1 i_2} \left(q_{ij}, a_{ij}, \theta \right)} \right]^{\mathbf{1}\left(s_{i_1 i_2} = -1\right)}, \tag{10}$$

where $Q_{ij} = \left(D_{ij0}, D_{ij3}, R_{ij0}, R_{ij1} \right)'$, $S_{ij} = D_{ij2} - D_{ij1}$ and

$$\begin{aligned} b^{01}_{ij} \left(q_{ij}, a_{ij}, \theta \right) &\overset{def}{\equiv} \frac{1 - F \left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij} \right)}{F \left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij} \right)} \frac{F \left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij} \right)}{1 - F \left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij} \right)} \\ b^{10}_{ij} \left(q_{ij}, a_{ij}, \theta \right) &\overset{def}{\equiv} \frac{F \left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij} \right)}{1 - F \left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij} \right)} \frac{1 - F \left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij} \right)}{F \left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij} \right)}. \end{aligned}$$

Observe that the denominator in (9) is a summation over $2^{\mathbf{m}_N}$ elements, where $\mathbf{m}_N$ is the number of stable dyads in the network. The remarkable feature of Theorem 1 is that this sum, unlike in many other similar contexts (e.g., Blitzstein and Diaconis, 2011; Chatterjee and Diaconis, 2013), is not intractable. Indeed the ratio (9) can be expressed as a simple product of just $\mathbf{m}_N$ terms. Given the interdependencies across dyads embedded in the (unconditional) likelihood, that such a factorization is possible hinges critically on the choice of conditioning set. The proof of this result, which is given below, requires careful bookkeeping and some counting/permutation arguments.

Two corollaries follow from Theorem 1 directly. The first is a semiparametric maximum-score type identification result.

Corollary 1. *(*SEMIPARAMETRIC IDENTIFICATION*) Under the data generating process specified in Section 1*

$$\Pr\left(D_{ij1}=0, D_{ij2}=1 \middle| Q_{ij}=q, Z_{ij}=1\right) - \Pr\left(D_{ij1}=1, D_{ij2}=0 \middle| Q=q, Z_{ij}=1\right) \lesseqqgtr 0$$

according to whether

$$\beta\left(d_3-d_0\right)+\gamma\left(r_1-r_0\right) \lesseqqgtr 0.$$

Proof. Follows from Theorem 1 and Manski (1987) or Honoré and Kyriazidou (2000). $\square$

Because R_{ijt} is integer-valued semiparametric point identification of $\theta=(\beta,\gamma)'$ is not possible (even after normalization). However if $R_{ij1}-R_{ij0}$ has a large number of support points, the identified set will be quite small.

The second corollary to Theorem 1 shows that point identification (up to scale) is possible under logistic errors.

Corollary 2. *(*LOGISTIC IDENTIFICATION*) When, additionally, U_{ijt} is standard logistic*

$$\Pr\left(D_{ij1}=d_1, D_{ij2}=d_2 \middle| Q_{ij}=q, Z_{ij}=1\right) = \left(\frac{\exp\left(x'\theta\right)}{1+\exp\left(x'\theta\right)}\right)^{\mathbf{1}(s=1)}\left(\frac{\exp\left(x'\theta\right)}{1+\exp\left(x'\theta\right)}\right)^{\mathbf{1}(s=-1)}$$

with $X_{ij}=\left(D_{ij3}-D_{ij0}, R_{ij1}-R_{ij0}\right)'$.

Proof. Follows from Theorem 1 and direct calculation. $\square$

As in the single agent dynamic binary choice case (e.g., Cox, 1958; Chamberlain, 1985; Honoré and Kyriazidou, 2000), Theorem 1 follows from an implication of the model that is invariant to the value of $\mathbf{A}$. In the single agent case a comparison of the *relative* frequencies of the link sequences $d_{ij0}01d_{ij3}$ and $d_{ij0}10d_{ij3}$ provides information about the strength of state-dependence. Consequently all agents who revise their choice between periods $t=1$ and $t=2$ contribute.

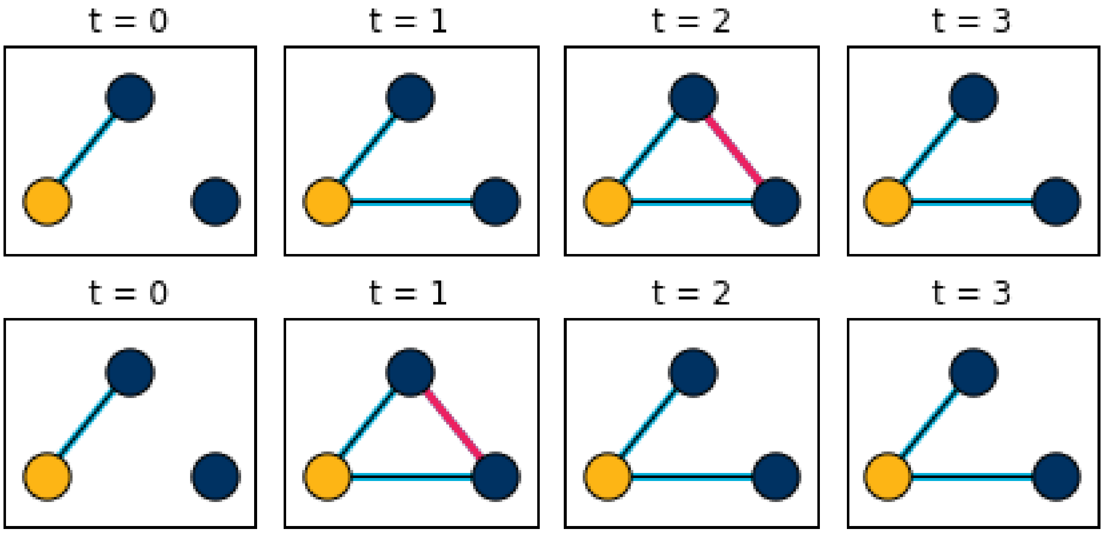
In the present context, the *relative* frequencies of the link sequences $d_{ij0}01d_{ij3}$ and $d_{ij0}10d_{ij3}$ also provides information about the signs and magnitudes of β_0 and γ_0. However we must confine analysis to dyads who, in addition to revising their linking decision between periods $t = 1$ and $t = 2$, are also embedded in stable neighborhoods. We specifically learn about β_0 versus γ_0 from the link histories of dyads embedded in stable neighborhoods with different types of network architecture. The need to condition on neighborhood stability arises because of interdependencies in link decisions.

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 the ij dyad). When linking decisions are interdependent, however, these two likelihoods may have many terms different, even though the two network sequences are nearly identical. That two nearly identical network sequences may have very different likelihoods attached to them 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 in value 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 links to form in period $t = 2$. Finally, the change in d_{ij1} does not affect 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.

The insight of Theorem 1 is that we can control these cascading effects on the likelihood by restricting how the neighborhood surrounding ij evolves. If ij is embedded in a stable neighborhood then a permutation of d_{ij1} and d_{ij2} will leave the *net* contribution of all non ij pairs to the likelihood unchanged. Specifically while many terms in the two likelihoods will be nominally different, it turns out that after permuting terms they can be shown to be identical up to the contribution from ij alone. The balance of the argument then follows from ideas introduced in the study of single agent discrete choice models (Chamberlain, 1985; Manski, 1987; Honoré and Kyriazidou, 2000).

Figure 3: Identification: transitivity versus homophily



Notes: Number agents 1, 2 and 3 clockwise from the top in each network. The top and bottom rows depict two network sequences. In the top one agents 1 and 2 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 1 and 2 constitute a stable dyad (since $d_{131} = d_{132} = d_{133} = 1$ and $d_{231} = d_{232} = d_{233} = 1$) and (ii) while they share a link in common in period $t = 1$, they do not in period $t = 0$. Consequently forming a link has a higher return in period 2 than in period 1. In period 2 the link generates utility from ensuring 'triadic closure', no such utility gain is generated by a period 1 link. Therefore, the top network sequence arises more frequently than the bottom in the presence of a structural taste for transitivity in links.

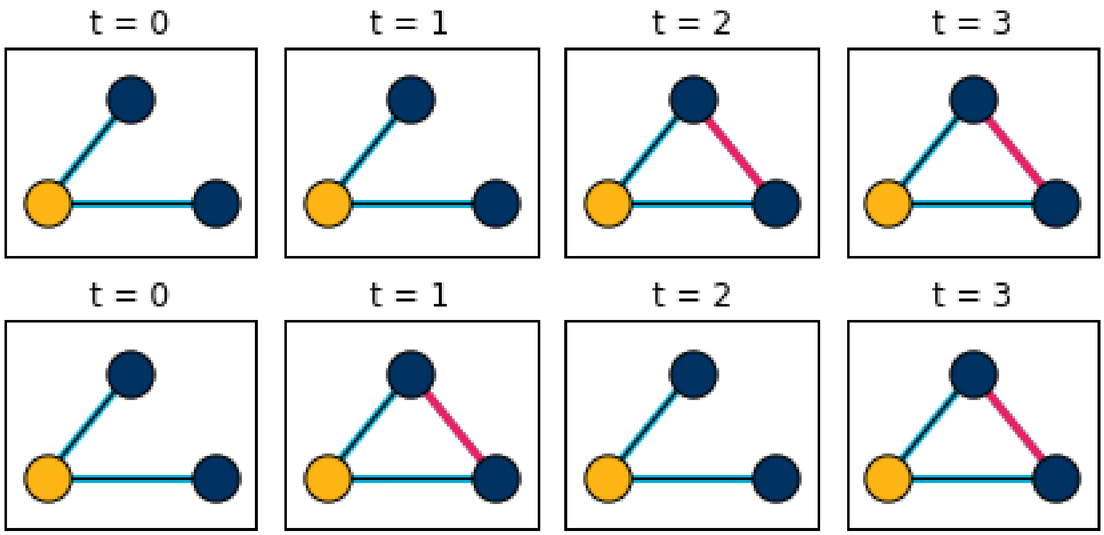
Examples of stable neighborhood

Before presenting the proof of Theorem 1 it is helpful to review a view examples illustrating how Corollary 1 works in practice.

The two rows in Figure 3 depict two network sequences. Numbering agents 1, 2 and 3 clockwise from the top, we can see that agents 1 and 2 (the two Berkeley blue nodes) constitute a stable dyad. Further observe that while these two agents share agent 3 (the California gold node) as a common friend in period $t = 1$, they do not in period $t = 0$. Therefore the returns to linking in period $t = 2$, where agents 1 and 2 reap the returns from engineering triadic closure, are higher than the corresponding returns from linking in period $t = 1$. In the presence of a structural taste for transitivity, $\gamma > 0$, we will observe the top sequence more frequently than the bottom sequence.

Figure 4 develops an example of how the relative frequency of two different network sequences provides information about the state dependence parameter β_0. Here the intuition parallels that familiar from the single agent binary choice case (e.g., Cox, 1958; Heckman, 1978; Chamberlain, 1985).

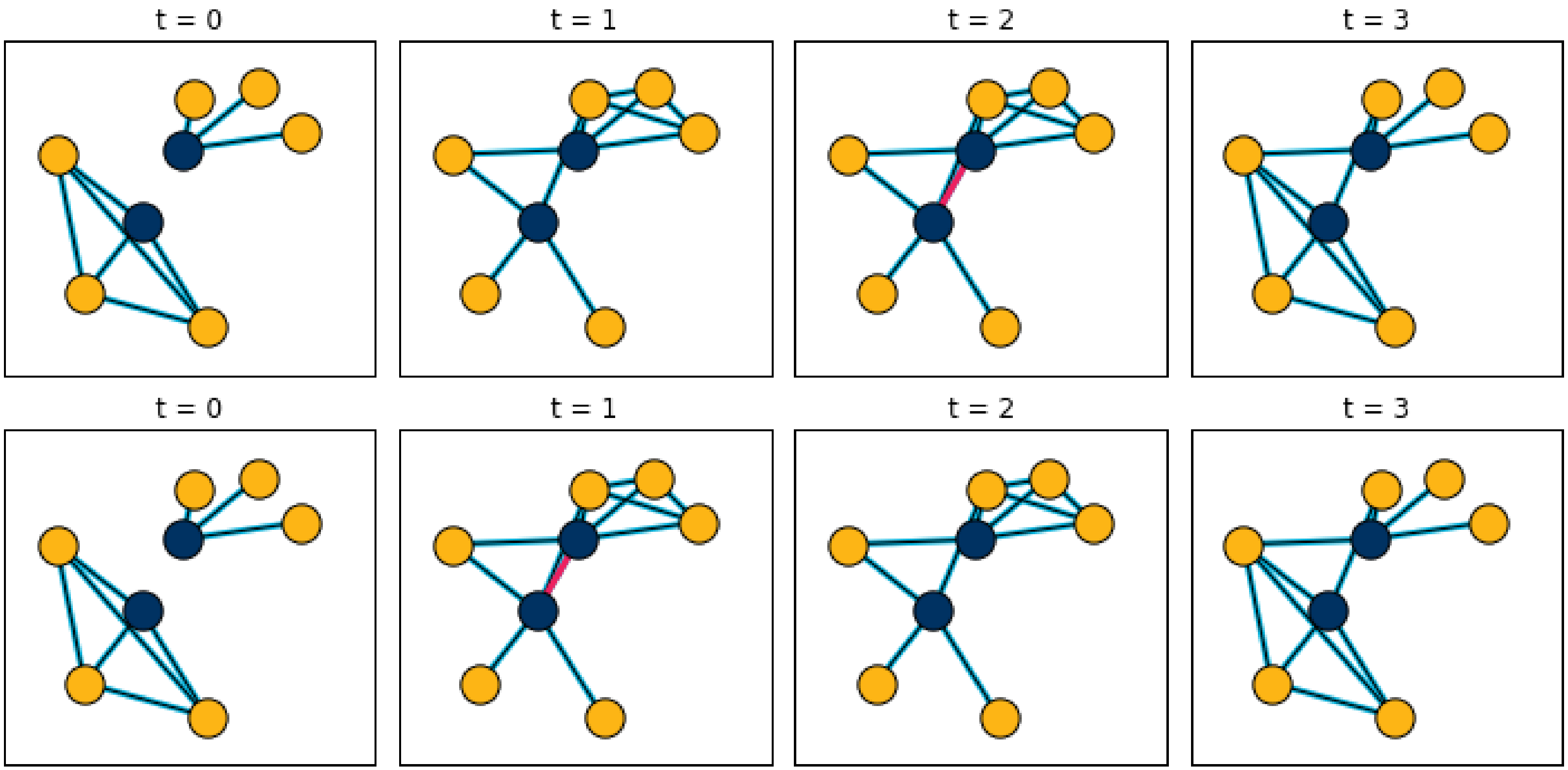
Figure 4: Identification: state dependence versus heterogeneity



Notes: See the notes to Figure 3. In this example $r_{120} = r_{121}$, so agents 1 and 2 will accrue returns from transitivity by linking in both periods 1 and 2. However, $d_{123} = 1$ and $d_{120} = 0$, suggesting that, in the presence of state dependence ($\beta_0 > 0$), the top sequence will occur more frequently than the bottom. The intuition in this case is very similar to that underlying the results of Cox (1958), Heckman (1978) and Chamberlain (1985).

As a final example consider the two network sequences depicted in Figure 5. The two Berkeley blue nodes constitute a stable dyad. Both blue nodes maintain the same links in periods $t = 1, 2, 3$ (except with each other) and it is also the case that the links maintained by their friends remain constant in periods $t = 1, 2$. Since the two blue nodes are neither linked in the initial $t = 0$ period or the final $t = 3$ period, state dependence does not play a role in their linking decisions. However, observe that the two agents share no links in common in period $t = 0$, while sharing two in common in period $t = 1$. Consequently the returns to the two agents linking in period $t = 2$ will be higher than those in available in $t = 1$. Consequently we should observe the top sequence more frequently than the bottom sequence in the presence of a structural taste for transitivity.

Figure 5: Identification: transitivity versus homophily in a more complex stable neighborhood



Notes: The two Berkeley blue nodes constitute a stable dyad in both the top and bottom sequences. Both blue nodes maintain the same links in periods $t = 1, 2, 3$ (except with each other) and it is also the case that the links maintained by their friends remain constant in periods $t = 1, 2$. See text for additional narrative.

Proof of Theorem 1

Readers uninterested in the proof of Theorem 1 can skip directly to Section 3, which discusses estimation and inference.

The proof of Theorem 1 relies on the following Lemma.

Lemma 2. *(*PERMUTATION*) Consider the network sequence formed by permuting the period 1 and 2 link decisions of any subset of the set of all stable dyads, $\mathcal{D}_s$. Let (R^*_{il1}, R^*_{il2}) denote the values of (R_{il1}, R_{il2}) after such a permutation for $i \in \mathcal{N}_s$ and $l \in \mathcal{N}^c_s$. Let ij be the stable dyad to which i belongs. If ij's link histories where changed as part of the permutation, then* $(R^*_{il1}, R^*_{il2}) = (R_{il2}, R_{il1})$, otherwise $(R^*_{il1}, R^*_{il2}) = (R_{il1}, R_{il2})$.

Proof. From the definition of R_{ilt} we have, for all $l = 1, \ldots, N$ not equal to i or j,

$$\begin{aligned} R_{il2} - R_{il1} &= \left[\sum_{k=1}^{N} D_{ik2} D_{lk2}\right] - \left[\sum_{k=1}^{N} D_{ik1} D_{lk1}\right] \\ &= \sum_{k=1, k\neq j}^{N} D_{ik1}\left(D_{lk2} - D_{lk1}\right) + D_{ij2}D_{lj2} - D_{ij1}D_{lj1} \\ &= \sum_{k=1, k\neq j}^{N} D_{ik1}\left(D_{lk2} - D_{lk1}\right) + \left(D_{ij2} - D_{ij1}\right) D_{lj1} \end{aligned}$$

with the second equality coming from rearrangement and the stability of all of i's links, other than those with j, across periods 1 and 2 (which implies $D_{ik1} = D_{ik2}$ for all $k \neq j$). The third equality comes from the corresponding stability of j's links (which implies $D_{lj2} = D_{lj1}$).

Now observe that $\sum_{k=1,k\neq j}^{N} D_{ik1}\left(D_{lk2} - D_{lk1}\right)$, the first term to the right of the last equality above, equals zero since if $D_{ik1} = 1$, then k is a neighbor of i and the definition of neighborhood stability then implies that $D_{lk1} = D_{lk2}$. Rearranging we have shown that

$$R_{il2} = R_{il1} + \left(D_{ij2} - D_{ij1}\right) D_{lj1}. \tag{11}$$

Now consider R^*_{il1}, the number of friends in common between agents i and l after permuting ij's period 1 and 2 link statuses:

$$\begin{aligned} R^*_{il1} &= R_{il1} + D_{ij2}D_{lj1} - D_{ij1}D_{lj1} \\ &= R_{il1} + \left(D_{ij2} - D_{ij1}\right) D_{lj1} \end{aligned}$$

which coincides with R_{il2} by (11) above. A similar argument gives $R^*_{il2} = R_{il1}$. □

To show the main result it is convenient to partition the likelihood (6) into four components:

$$\begin{aligned}
p\left(\mathbf{d}_0^3,\mathbf{a},\theta\right) &= \pi\left(\mathbf{d}_0,\mathbf{a}\right) && (12)\\
&\times \prod_{\mathbf{i}\in\mathcal{D}_{od}}\prod_{t=1}^{3} F\left(\beta d_{i_1i_2t-1}+\gamma r_{i_1i_2t-1}+a_{i_1i_2}\right)^{d_{i_1i_2t}} && (13)\\
&\times\left[1-F\left(\beta d_{i_1i_2t-1}+\gamma r_{i_1i_2t-1}+a_{i_1i_2}\right)\right]^{1-d_{i_1i_2t}} &&\\
&\times \prod_{\mathbf{i}\in\mathcal{D}_{ns}}\prod_{t=1}^{3} F\left(\beta d_{i_1i_2t-1}+\gamma r_{i_1i_2t-1}+a_{i_1i_2}\right)^{d_{i_1i_2t}} && (14)\\
&\times\left[1-F\left(\beta d_{i_1i_2t-1}+\gamma r_{i_1i_2t-1}+a_{i_1i_2}\right)\right]^{1-d_{i_1i_2t}} &&\\
&\times \prod_{\mathbf{i}\in\mathcal{D}_{s}}\prod_{t=1}^{3} F\left(\beta d_{i_1i_2t-1}+\gamma r_{i_1i_2t-1}+a_{i_1i_2}\right)^{d_{i_1i_2t}} && (15)\\
&\times\left[1-F\left(\beta d_{i_1i_2t-1}+\gamma r_{i_1i_2t-1}+a_{i_1i_2}\right)\right]^{1-d_{i_1i_2t}} &&
\end{aligned}$$

Line (12) contains the likelihood contribution corresponding to the joint density of the initial network condition and pair-specific heterogeneity terms. Line (13) contains the contributions of all dyads where neither agent is embedded in a stable neighborhood. Line (14) contains the contribution from all dyads where one agent, but not the other, is embedded in a stable neighborhood. Finally line (15) contains the contribution of dyads embedded in stable neighborhoods.

We can use this partition of the likelihood to evaluate the conditional likelihood of observing $\mathbf{D}_0^3=\mathbf{d}_0^3$ conditional on the event that $\mathbf{d}_0^3\in\mathbb{V}^s$

$$l^c\left(\mathbf{d}_0^3,\mathbf{a},\theta\right) = \frac{p\left(\mathbf{d}_0^3,\mathbf{a},\theta\right)}{\sum_{\mathbf{v}\in\mathbb{V}^s}p\left(\mathbf{v}_0^3,\mathbf{a},\theta\right)}. \tag{16}$$

In order to simplify (16) it is convenient to analyze its inverse , which consists of the sum of $2^{\mathbf{m}_N}$ ratios of the form

$$\frac{p\left(\mathbf{v}_0^3,\mathbf{a},\theta\right)}{p\left(\mathbf{d}_0^3,\mathbf{a},\theta\right)}, \tag{17}$$

for $\mathbf{v}_0^3\in\mathbb{V}^s$. We can use the likelihood decomposition given by (12) to (15) above to derive an explicit expression for this ratio.

First, since the initial condition is held fixed across the two network sequences, the $\pi\left(\mathbf{v}_0,\mathbf{a}\right)$ and $\pi\left(\mathbf{d}_0,\mathbf{a}\right)$ terms in, respectively, the numerator and denominator of (17) cancel. Second, observe that any permutation of the period 1 and 2 link decisions of dyads in $\mathcal{D}_s$ leaves the utility associated with links across dyads in $\mathcal{D}_{od}$ unchanged. Therefore the period $t=1,2,3$ likelihood contributions associated with all dyads in $\mathcal{D}_{od}$, the terms in line (13) above, also cancel in (17).

Third, consider dyad il with $i \in \mathcal{N}_s$ and $l \in \mathcal{N}_s^c$. Let r^*_{il0}, r^*_{il1} and r^*_{il2} denote the number of friends i and l have in common in periods $t = 0, t = 1$ and $t = 2$, respectively, under network sequence $\mathbf{v}_0^3 \in \mathbb{V}^s$. Since i is part of a stable dyad (and l is not), it must be the case that $v_{il1} = v_{il2} = v_{il3} = 1$ or $v_{il1} = v_{il2} = v_{il3} = 0$ (by the first part of Definition 1). Note also that, by the definition of $\mathbb{V}^s$, $r^*_{il0} = r_{il0}$, $v_{il0} = d_{il0}$, $v_{il1} = d_{il1}$, $v_{il2} = d_{il2}$ and $v_{il3} = d_{il3}$. First assume that $v_{il1} = v_{il2} = v_{il3} = 1$. In this case the period $t = 1, 2, 3$ likelihood contributions of the il dyad under $\mathbf{v}_0^3$ are

$$
\begin{aligned}
F\left(\beta v_{il0} + \gamma r^*_{il0} + a_{il}\right) F\left(\beta + \gamma r^*_{il1} + a_{il}\right) F\left(\beta + \gamma r^*_{il2} + a_{il}\right) &= F\left(\beta d_{il0} + \gamma r_{il0} + a_{il}\right) \\
&\quad \times F\left(\beta + \gamma r_{il2} + a_{il}\right) \\
&\quad \times F\left(\beta + \gamma r_{il1} + a_{il}\right),
\end{aligned}
$$

where the equality follows from the fact that $r^*_{il1} = r_{il2}$ and $r^*_{il2} = r_{il1}$ by Lemma 2 above. Now observe that the expression to the right of the equality, after re-ordering, is identical to the $t = 1, 2, 3$ likelihood contributions of the il dyad under $\mathbf{d}_0^3$. Using an analogous observation for the $v_{il0} = v_{il2} = v_{il3} = 0$ case implies that the *net* period $t = 1, 2, 3$ contributions of dyad il to the likelihoods of $\mathbf{d}_0^3$ and $\mathbf{v}_0^3$ are identical (up to a re-ordering of terms). This this leads to a cancellation of the terms corresponding to lines (14) above. This is a key step of the argument.

All that remains of (17) is the ratio of the $\mathbf{m}_N$ components in line (15), one for each stable dyad, under $\mathbf{v}$ versus $\mathbf{d}$. If $(v_{ij1}, v_{ij2}) = (d_{ij1}, d_{ij2})$, then the contributions of dyad $ij \in \mathcal{D}_s$ to the numerator and denominator of (17) cancel. If $(v_{ij1}, v_{ij2}) = (d_{ij2}, d_{ij1}) = (1, 0)$, then dyad $ij \in \mathcal{D}_s$ contributes the term

$$
\begin{aligned}
b^{10}_{ij}\left(\mathbf{v}_0^3, \theta\right) &= \frac{F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)}{1 - F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)} \frac{1 - F\left(\beta + \gamma r_{ij1} + a_{ij}\right)}{F\left(\gamma r_{ij1} + a_{ij}\right)} \\
&\quad \times \frac{F\left(\gamma r_{ij2} + a_{ij}\right)^{d_{ij3}}}{F\left(\beta + \gamma r_{ij2} + a_{ij}\right)^{d_{ij3}}} \frac{\left[1 - F\left(\gamma r_{ij2} + a_{ij}\right)\right]^{1-d_{ij3}}}{\left[1 - F\left(\beta + \gamma r_{ij2} + a_{ij}\right)\right]^{1-d_{ij3}}} \\
&= \frac{F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)}{1 - F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)} \frac{1 - F\left(\beta + \gamma r_{ij1} + a_{ij}\right)}{F\left(\gamma r_{ij1} + a_{ij}\right)} \\
&\quad \times \frac{F\left(\gamma r_{ij1} + a_{ij}\right)^{d_{ij3}}}{F\left(\beta + \gamma r_{ij1} + a_{ij}\right)^{d_{ij3}}} \frac{\left[1 - F\left(\gamma r_{ij1} + a_{ij}\right)\right]^{1-d_{ij3}}}{\left[1 - F\left(\beta + \gamma r_{ij1} + a_{ij}\right)\right]^{1-d_{ij3}}},
\end{aligned}
$$

to (17), which, following Honoré and Kyriazidou (2000), can be shown to equal

$$
b^{10}_{ij}\left(\mathbf{v}_0^3, \theta\right) = \frac{F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)}{1 - F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)} \frac{1 - F\left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij}\right)}{F\left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij}\right)} \stackrel{def}{\equiv} b^{10}_{ij}\left(q_{ij}, a_{ij}, \theta\right) \quad (18)
$$

if $d_{ij3} = 1$ and also if $d_{ij3} = 0$. In the logit case

$$b_{ij}^{10}(q_{ij}, a_{ij}, \theta) = \exp\left(-\beta\left(d_{ij3} - d_{ij0}\right) - \gamma\left(r_{ij1} - r_{ij0}\right)\right),$$

which is invariant to the value of a_{ij}. Recall that q_{ij} denotes the conditioning vector $q_{ij} = (d_{ij0}, d_{ij3}, r_{ij0}, r_{ij1})'$ and $\theta = (\beta, \gamma)'$ the parameter vector of interest.

If $(v_{ij1}, v_{ij2}) = (d_{ij2}, d_{ij1}) = (0, 1)$, then dyad $ij \in \mathcal{D}_s$ contributes

$$\begin{aligned} b_{ij}^{01}\left(\mathbf{v}_0^3, \theta\right) &= \frac{1 - F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)}{F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)} \frac{F\left(\gamma r_{ij1} + a_{ij}\right)}{1 - F\left(\beta + \gamma r_{ij1} + a_{ij}\right)} \\ &\quad \times \frac{F\left(\beta + \gamma r_{ij2} + a_{ij}\right)^{d_{ij3}}}{F\left(\gamma r_{ij2} + a_{ij}\right)^{d_{ij3}}} \frac{\left[1 - F\left(\beta + \gamma r_{ij2} + a_{ij}\right)\right]^{1-d_{ij3}}}{\left[1 - F\left(\gamma r_{ij2} + a_{ij}\right)\right]^{1-d_{ij3}}} \\ &= \frac{1 - F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)}{F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)} \frac{F\left(\gamma r_{ij1} + a_{ij}\right)}{1 - F\left(\beta + \gamma r_{ij1} + a_{ij}\right)} \\ &\quad \times \frac{F\left(\beta + \gamma r_{ij1} + a_{ij}\right)^{d_{ij3}}}{F\left(\gamma r_{ij1} + a_{ij}\right)^{d_{ij3}}} \frac{\left[1 - F\left(\beta + \gamma r_{ij1} + a_{ij}\right)\right]^{1-d_{ij3}}}{\left[1 - F\left(\gamma r_{ij1} + a_{ij}\right)\right]^{1-d_{ij3}}} \end{aligned}$$

which, again follows following Honoré and Kyriazidou (2000), equals

$$b_{ij}^{01}\left(\mathbf{v_0^3}, \theta\right) = \frac{1 - F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)}{F\left(\beta d_{ij0} + \gamma r_{ij0} + a_{ij}\right)} \frac{F\left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij}\right)}{1 - F\left(\beta d_{ij3} + \gamma r_{ij1} + a_{ij}\right)} \overset{def}{\equiv} b_{ij}^{01}(q_{ij}, a_{ij}, \theta) \quad (19)$$

if $d_{ij3} = 1$ and also if $d_{ij3} = 0$. In the logit case

$$b_{ij}^{01}(q_{ij}, a_{ij}, \theta) = \exp\left(\beta\left(d_{ij3} - d_{ij0}\right) + \gamma\left(r_{ij1} - r_{ij0}\right)\right).$$

For all $\mathbf{v}_0^3 \in \mathbb{V}^s$ we therefore have that the contributions of dyad $ij \in \mathcal{D}_s$ to the numerator and denominator of (17) either cancel or equal (18) or (19) according to whether $(d_{ij1}, d_{ij2}) = (0, 1)$ or $(d_{ij1}, d_{ij2}) = (1, 0)$.

Let $S_{ij} = D_{ij2} - D_{ij1}$, using the above calculations, we can write,

$$\frac{\sum_{\mathbf{v} \in \mathbb{V}^s} p\left(\mathbf{v}_0^3, \mathbf{a}, \theta\right)}{p\left(\mathbf{d}_0^3, \mathbf{a}, \theta\right)} = \prod_{\mathbf{i} \in \mathcal{D}_s} \left[1 + b_{i_1 i_2}^{01}(q_{ij}, a_{ij}, \theta)\right]^{\mathbf{1}\left(s_{i_1 i_2} = 1\right)} \left[1 + b_{i_1 i_2}^{10}(q_{ij}, a_{ij}, \theta)\right]^{\mathbf{1}\left(s_{i_1 i_2} = -1\right)}.$$

Note that the product to the right of the equality above evaluates to a sum of $2^{\mathbf{m}_N}$ terms, one for each element of $\mathbb{V}^s$, as required. Inverting yields a simplification of (16) equal to (10) as claimed.

In the logit case (10) simplifies further to

$$l^{c}(\theta) = \prod_{\mathbf{i}\in\mathcal{D}_s}\left(\frac{\exp\left(x'_{i_1 i_2}\theta\right)}{1+\exp\left(x'_{i_1 i_2}\theta\right)}\right)^{\mathbf{1}\left(s_{i_1 i_2}=1\right)}\left(\frac{\exp\left(x'_{i_1 i_2}\theta\right)}{1+\exp\left(x'_{i_1 i_2}\theta\right)}\right)^{\mathbf{1}\left(s_{i_1 i_2}=-1\right)}, \tag{20}$$

recalling that $X_{ij} = (D_{ij3} - D_{ij0}, R_{ij1} - R_{ij0})'$.

3 Estimation and inference

In this section I introduce an estimator for θ_0 based on Corollary 2. This corresponds to the case where the model described in Section 1 is augmented with the assumption that U_{ijt} is a standard logistic random variable. The estimator is simply the maximizer of the conditional likelihood presented in Theorem 1 (evaluated under the logit assumption):

$$L_N(\theta) = \binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i} l_{ij}(\theta). \tag{21}$$

where

$$l_{ij}(\theta) = Z_{ij}\left\{S_{ij}X'_{ij}\theta - \ln\left[1+\exp\left(S_{ij}X'_{ij}\theta\right)\right]\right\}. \tag{22}$$

The *stable neighborhood logit* (henceforth SN logit) estimate of θ_0 is given by[9]

$$\hat{\theta}_{\mathrm{SN}} = \arg\max_{\theta\in\Theta} L_N(\theta) \tag{24}$$

To derive the large sample properties of $\hat{\theta}_{\mathrm{SN}}$ it is helpful to first formalize some conditional independence properties embedded in the conditional likelihood (10). Order dyads, without loss of generality, so that $\mathbf{i} = \mathbf{1}, \ldots, \mathbf{m}_N$ indexes the $\mathbf{m}_N = |\mathcal{D}_s|$ dyads embedded in stable neighborhoods.

Lemma 3. (Independence of Link Histories Among Stable Dyads) The events $S_{\mathbf{1}} = s, \ldots, S_{\mathbf{m}_N} = s$ are conditionally independent given $Z_{\mathbf{1}} = \cdots = Z_{\mathbf{m}_N} = 1$, $\mathbf{D}_0 = \mathbf{d}_0$, $\mathbf{D}_3 = \mathbf{d}_3$ and $\mathbf{A} = \mathbf{a}$.

[9] While I will not present formal results for it, a semiparametric maximum score estimator based on Theorem 1 could be constructed by maximizing

$$\max_{\theta:\|\theta'\theta\|=1}\binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i} Z_{ij}\left(D_{ij2} - D_{ij1}\right)\operatorname{sgn}\left\{X'_{ij}\theta\right\}. \tag{23}$$

The set of maximizers of (23) would provide an estimate of the identified set.

Proof. The conditional likelihood (10) evaluated at $\theta = \theta_0$ gives the probability of all possible permutations of period 1 and period 2 link histories between dyads embedded in stable neighborhoods conditional on all other features of the network being held fixed. For any $\mathbf{i} = \mathbf{1}, \ldots, \mathbf{m}_N$ marginalizing gives

$$\begin{aligned} \Pr\left(S_{\mathbf{i}} = s \middle| Z_{\mathbf{1}} = \cdots = Z_{\mathbf{m}} = 1, \mathbf{d}_0, \mathbf{d}_3, \mathbf{a}\right) &= \left(\frac{1}{1 + b_{i_1 i_2}^{01}\left(q_{ij}, a_{ij}, \theta_0\right)}\right)^{\mathbf{1}(s_{\mathbf{i}}=1)} \\ &\quad \times \left(\frac{1}{1 + b_{i_1 i_2}^{10}\left(q_{ij}, a_{ij}, \theta_0\right)}\right)^{\mathbf{1}(s_{\mathbf{i}}=-1)} \\ &= \Pr\left(S_{\mathbf{i}} = s \middle| Z_{\mathbf{i}} = 1, q_{\mathbf{i}}, a_{\mathbf{i}}\right) \end{aligned} \tag{25}$$

the product of (25) for $\mathbf{i} = \mathbf{1}, \ldots, \mathbf{m}_N$ gives (20). $\square$

Lemma 3 implies that the stable neighborhood logit criterion function (20) consists of $\mathbf{m}_N = |\mathcal{D}_s|$ conditionally independent components. As long as $\mathbf{m}_N \to \infty$ and $N \to \infty$ consistency and asymptotic normality of $\hat{\theta}_{\mathrm{SN}}$ follows relatively directly. To state a formal result I require some additional assumptions.

Assumption 1. (Sampling) *The econometrician observes all agents and links between them in* $t = 0, 1, 2, 3$.

Developing results for other sampling schemes, as when the researcher collects a set of ego-centered graphs, would be an interesting topic for future search.

Assumption 2. (Regularity)
(i) $\theta_0 \in \operatorname{int}(\Theta)$, *with* Θ *a compact subset of* R^2,
(ii) for any two agents i *and* j, R_{ijt} *is a bounded for* $t = 0, 1$

Part (i) of Assumption 2 is standard regularity condition. Given that most networks are sparse, part (ii) is not especially restrictive. Specifically if each agent can maintain only a finite number of links, then she can have at most a finite number of links in common with any other agent. This assumption can also be relaxed by assuming that R_{ijt} has, instead, a sufficient number of finite moments.

Point identification further requires:

Assumption 3. (Identification)
(i) $n\alpha_N \to \infty$ *with* $\alpha_N = \Pr\left(Z_{ij} = 1\right)$
(ii) $\mathbb{E}\left[X_{ij}X_{ij}' \middle| Z_{ij} = 1\right]$ *is a finite non-singular matrix*

Part (i) of Assumption 3 ensures that the number of stable dyads increases with the size of the network. Although it allows the probability attached to the event "ij is a stable dyad" to approach zero as the network grows, it does so at sufficiently slow rate such that the total number of stable dyads nevertheless grows large as $N \to \infty$. One way to ensure that part (i) holds is to impose restrictions on the sequence $\{A_{ij}\}_{j=i+1}^{\infty}$. An example is implicitly provided by the Monte Carlo design introduced below. The single agent binary choice analog of part (i) of Assumption 3 is the requirement that the number of switchers or movers grows with the sample size (Chamberlain, 1980).

Part (ii) of Assumption 3 is a standard identification condition for binary choice models, albeit expressed conditionally on $Z_{ij} = 1$ (e.g., Amemiya, 1985).

Let $p_{ij}(\theta) = \frac{\exp\left(X_{ij}'\theta\right)}{1+\exp\left(X_{ij}'\theta\right)}$ with $p_{ij} = p_{ij}(\theta_0)$ and $\Lambda_0 = \mathbb{E}\left[p_{ij}(1-p_{ij}) X_{ij}X_{ij}' \middle| Z_{ij} = 1\right]$. The large sample properties of $\hat{\theta}_{\text{SN}}$ are summarized in the following theorem.

Theorem 2. (Asymptotic Properties of $\hat{\theta}_{\text{SN}}$) *Under the assumptions stated above*
(i) $\hat{\theta}_{\text{SN}} \xrightarrow{p} \theta_0$
(ii) $\sqrt{n\alpha_N}\left(\hat{\theta}_{\text{SN}} - \theta_0\right) \xrightarrow{D} N\left(0, \Lambda_0^{-1}\right)$

The proof of Theorem 2 can be found in Appendix A. It is relatively straightforward. Recall that $\mathbf{m}_N$ equals the number of conditionally independent components in the SN logit criterion function. Therefore $\{\mathbf{m}_N\}$ is a sequence of integer-valued random variables. Part (i) of Assumption 3 implies that

$$\frac{\mathbf{m}_N}{n\alpha_N} \xrightarrow{p} 1,$$

with $n\alpha_N \to \infty$. Hence the number of components in the SN logit criterion function will grow large with the network. This allows for the application of a central limit theorem for a random number of summands. Theorem 1.9.4 of Serfling (1980) is sufficient for my purposes.

The upshot of Theorem 2 is that, in practice, estimation of, and inference on, θ_0 is very straightforward. The main challenge is to quickly find all stable dyads in the network. As mentioned above, a description of how to construct the Z_{ij} indicator can be found in Appendix B and a Python function for finding stable dyads is included in the supplementary materials.

Once all stable dyads have been located, estimation and inference can be conducted using a standard logit maximum likelihood program. Specifically let $Y_{ij} = \mathbf{1}(S_{ij} = 1) = \mathbf{1}(D_{ij1} = 0, D_{ij1} = 1)$ and recall that $X_{ij} = (D_{ij3} - D_{ij0}, R_{ij1} - R_{ij0})'$. A standard logit fit of Y_{ij} onto X_{ij} across the subset of $Z_{ij} = 1$ stable dyads coincides with the SN logit estimator. The conventional standard errors reported by the program will be asymptotically valid.[10]

[10] A constant should not be included in the design matrix. Most logit programs include a constant by

4 Monte Carlo experiments

The Monte Carlo design uses a random geometric graph to construct an "opportunity graph" for link formation. 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

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

with U_{ij0} logistic and A_{ij0} taking one of two values. If the Euclidean distance between i and j is less than or equal to r, then $A_{ij0} = \ln\left(\frac{0.75}{1-0.75}\right)$, otherwise A_{ij0} 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 I can manipulate the average degree, and hence connectivity, of the initial condition. I choose values of r such that in large networks average degree in period $t = 0$ is 2, 3 or 4.[11] 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 almost all 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 (2) with $\beta = \gamma = 1$.[12] 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

default, but also have an exclusion option.

[11]As 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.

[12]Since agents greater that 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.

for triadic closure. Link clustering in these design 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 2.

A typical sequence of networks, with $N = 200$ and average degree in the initial network set equal to 4, is depicted in Figure 6. The figure indicates that stable dyads come in different forms and can arise even in a strongly connected network. In the figure the larger Berkeley blue nodes correspond to stable dyads, the California gold nodes are immediate neighbors, while the Medalist colored nodes are indirect neighbors. Stable dyads arise in sparse regions of the network, as might be expected, but also in relatively dense regions.

Table 1: Basic Properties of Simulated Network Sequences

Asymptotic Degree	2			3			4		
Period	$\mathbb{E}[D_{it}]$	T	GC	$\mathbb{E}[D_{it}]$	T	GC	$\mathbb{E}[D_{it}]$	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	4.98	0.58	0.83
$t = 2$	2.49	0.59	0.01	3.80	0.59	0.09	5.12	0.59	0.84
$t = 3$	2.50	0.60	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 $\mathbb{E}[D_{it}]$ 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 $\hat{\beta}$ and $\hat{\gamma}$ across the different designs. In all cases I set the number of agents equal to $N = 5,000$ and complete $B = 1,000$ Monte Carlo replications. The network size was chosen through trial and error to ensure the presence of a sufficient number of identifying stable dyads. With five thousand agents the number of stable dyads averages between 100 and 250 for the designs considered here.

In all cases the SN logit estimator is approximately mean and median unbiased and the associated Wald-based confidence intervals have actual coverage close to nominal 95 percent coverage.

In empirical work it is common to fit simple dynamic logistic regression models to network data. In this approach current linking decisions are modeled as a function of past network structure (e.g., Gulati and Gargiulo, 1999; Kossinets and Watts, 2009; Almquist and Butts, 2013). In the Monte Carlo designs considered here, this approach resulted in highly biased point estimates and nominal 95 percent confidence intervals with zero coverage in all cases (results not reported).[13]

[13]With $B = 1,000$ Monte Carlo replications the standard error of the coverage estimates are $\sqrt{0.95\,(1-0.95)\,/1,000} \approx 0.007$. Hence the null of actual coverage equaling 0.95 is accepted in all cases.

Table 2: Sampling properties of $\hat{\beta}$ and $\hat{\gamma}$

Asymptotic Degree	2		3		4	
$N = 5,000$	β	γ	β	γ	β	γ
Mean	1.0138	1.0458	1.0314	1.0421	1.0438	1.0456
Median	1.0018	1.0187	1.0061	1.0064	1.0410	1.0133
Std. Dev.	0.2661	0.2888	0.3575	0.2832	0.4575	0.2976
Mean Std. Err.	0.2593	0.2791	0.3375	0.2717	0.4493	0.2917
Coverage	0.9600	0.9530	0.9480	0.9600	0.9620	0.9650
Avg. # of Stable Dyads	237.5		162.9		110.6	
# of cvg. failures	0		0		1	

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

5 Conclusion

This paper has introduced a simple model of dynamic network formation which incorporates, for the first time, both a structural taste for transitivity and arbitrary homophily on time-invariant agent attributes. Transitivity and homophily are the two most widely posited reasons for link clustering in real world networks (e.g., Snijders, 2011, 2013). The model introduced here provides a means to assess the roles played by these two forces in practice. Taste transitivity induces a strong dependence in link formation trajectories across agents. By focusing on link histories across dyads embedded in stable neighborhoods, these dependencies can be controlled, and positive identification results derived. Estimation and inference is surprisingly simple, requiring a routine to find stable dyads, and a standard logit maximum likelihood program.

Several different directions for future research suggest themselves. First, the application of the methods proposed here to real world network data would be interesting. Second, it is an open question whether the notion of neighborhood stability can be extended to accommodate more complex surplus functions. In particular those with additional network interdependencies beyond a taste for transitivity. Third, the fixed effects results presented here provide justification for the further exploration of random effects approaches to dynamic network analysis.

A Proof of Theorem 2

This appendix presents a derivation of the large sample properties of the SN logit estimator summarized in Theorem 2 of the main text. All notation is as defined in the main text unless stated otherwise. The abbreviation TI refers to the Triangle Inequality, CSI to the Cauchy-Schwartz Inequality, LLN to the Law of Large Numbers and CLT to the Central Limit Theorem.

Consistency

To show part (i) of Theorem 2 consider the normalized SN logit criterion function

$$\alpha_N^{-1} L_N(\theta) = \alpha_N^{-1} \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} l_{ij}(\theta). \tag{26}$$

A dominating function for the "kernel" of (26) can be constructed as follows

$$\begin{aligned}
\left|\alpha_N^{-1} l_{ij}(\theta)\right| &= \alpha_N^{-1} Z_{ij} \left| -\ln 2 + \left(S_{ij} X_{ij}' - \frac{\exp\left(S_{ij} X_{ij}' \bar{\theta}\right)}{1+\exp\left(S_{ij} X_{ij}' \bar{\theta}\right)} S_{ij} X_{ij}' \right) \theta \right| \\
&\leq \alpha_N^{-1} Z_{ij} \left|\ln 2\right| + \alpha_N^{-1} Z_{ij} \left(\frac{1}{1+\exp\left(S_{ij} X_{ij}' \bar{\theta}\right)} \right) \left| S_{ij} X_{ij}' \theta \right| \\
&\leq \alpha_N^{-1} Z_{ij} \left|\ln 2\right| + \alpha_N^{-1} Z_{ij} \left| S_{ij} X_{ij}' \theta \right| \\
&= \alpha_N^{-1} Z_{ij} \left|\ln 2\right| + \alpha_N^{-1} Z_{ij} \left| X_{ij}' \theta \right| \\
&\leq \alpha_N^{-1} Z_{ij} \left|\ln 2\right| + \alpha_N^{-1} Z_{ij} \left\| X_{ij} \right\| \left\| \theta \right\|.
\end{aligned} \tag{27}$$

The first equality follows from a mean value expansion in θ, the second inequality by the TI, the third inequality by the fact that $[1+\exp(v)]^{-1}$ lies between zero and one, the fourth equality by the fact that $|S_{ij}| = 1$ if $Z_{ij} = 1$ and zero otherwise, and the last inequality by the CSI. By Assumption 2 the term to the right of the final inequality is bounded.

Taking expectations of both side of (27) therefore yields

$$\mathbb{E}\left[\left|\alpha_N^{-1} l_{ij}(\theta)\right|\right] \leq |\ln 2| + \mathbb{E}\left[\left\|X_{ij}\right\| \middle| Z_{ij} = 1\right] \|\theta\| = O(1),$$

which allows for an application of a law of large numbers for U-statistics with sample-size dependent kernels (e.g., Lemma A.3 of Ahn and Powell (1993)). This gives $\alpha_N^{-1} L_N(\theta) \xrightarrow{p}$

$Q(\theta)$ for

$$\begin{aligned} Q(\theta) &= \mathbb{E}\left[\alpha_N^{-1} l_{ij}(\theta)\right] \\ &= -\alpha_N^{-1} \Pr(Z_{ij}=1)\left\{\mathbb{E}\left[D_{\mathrm{KL}}(p_{ij} \| p_{ij}(\theta)) | Z_{ij}=1\right] + \mathbb{E}\left[\mathbf{S}(p_{ij}) | Z_{ij}=1\right]\right\} \\ &= -\mathbb{E}\left[D_{\mathrm{KL}}(p_{ij} \| p_{ij}(\theta)) | Z_{ij}=1\right] + \mathbb{E}\left[\mathbf{S}(p_{ij}) | Z_{ij}=1\right]. \end{aligned} \tag{28}$$

Here $D_{\mathrm{KL}}(p_{ij} \| p_{ij}(\theta))$ equals the Kullback-Liebler divergence of $p_{ij}(\theta) = \frac{\exp(X_{ij}'\theta)}{1+\exp(X_{ij}'\theta)}$ from $p_{ij} = p_{ij}(\theta_0)$ and $\mathbf{S}(p_{ij})$ is the binary entropy function. Clearly θ_0 is a maximizer of (28); part (ii) of Assumption 3 implies that θ_0 is the unique maximizer. By concavity of $L_N(\theta)$ in θ the convergence of $\alpha_N^{-1} L_N(\theta)$ to $Q(\theta)$ is uniform in $\theta \in \Theta$. Since conditions A, B and C of Theorem 4.1.1 of Amemiya (1985) hold, part (i) of the Theorem follows.

Asymptotic normality

To show part (ii) of Theorem 2 I begin with a mean value expansion of the first order condition of the (normalized) SN logit criterion function (26). This gives, after some re-arrangement,

$$\begin{aligned} \hat{\theta} - \theta_0 &= -\left[\alpha_N^{-1} \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} \nabla_{\theta\theta} l_{ij}(\bar{\theta})\right]^{+} \\ &\quad \times \left[\alpha_N^{-1} \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} \nabla_{\theta} l_{ij}(\theta_0)\right], \end{aligned} \tag{29}$$

where $\bar{\theta}$ lies "between" $\hat{\theta}$ and θ_0, $+$ denotes a Moore-Penrose generalized inverse and

$$\begin{aligned} \nabla_{\theta} l_{ij}(\theta) &= Z_{ij}\left\{\mathbf{1}(S_{ij}=1) - \frac{\exp(X_{ij}'\theta)}{1+\exp(X_{ij}'\theta)}\right\} X_{ij} \\ \nabla_{\theta\theta} l_{ij}(\theta) &= -Z_{ij}\left\{\frac{\exp(X_{ij}'\theta)}{\left[1+\exp(X_{ij}'\theta)\right]^2}\right\} X_{ij} X_{ij}'. \end{aligned}$$

By Assumption 2 and part (ii) of Assumption 3 $\mathbb{E}\left[\sup_{\theta\in\Theta} \|\nabla_{\theta\theta} l_{ij}(\theta)\|_{\max}\right] < \infty$ and continuous in $\theta \in \Theta$ so that, using part (i) of Assumption 3, $\alpha_N^{-1}\binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} \nabla_{\theta\theta} l_{ij}(\bar{\theta}) \xrightarrow{p} \Lambda_0$ with Λ_0 as defined immediately before the statement of Theorem 2 in the main text. The matrix Λ_0 is finite and non-singular by part (ii) of Assumption 3.

Asymptotic normality follows by demonstrating that the second term in (29), suitably normalized, obeys a central limit theorem (CLT). To show this, start by calculating the variance

$$\begin{aligned}
\mathbb{E}\left[\nabla_\theta l_{ij}\left(\theta_0\right)\nabla_\theta l_{ij}\left(\theta_0\right)\right] &= \Pr\left(Z_{ij}=1\right)\mathbb{E}\left[\left. q_{ij}\left(1-q_{ij}\right)X_{ij}X_{ij}'\right| Z_{ij}=1\right] \\
&= \alpha_N\Lambda_0 \\
&= O\left(\alpha_N\right).
\end{aligned}$$

Note also that $\mathbb{E}\left[\nabla_\theta l_{ij}\left(\theta_0\right)\nabla_\theta l_{ik}\left(\theta_0\right)\right]$ identically equals zero for any $j \neq k$ because each agent can belong to *at most* one stable dyad (so that $Z_{ij}Z_{ik}=0$). This observation, and a Hoeffding (1948) decomposition, gives

$$\begin{aligned}
\mathbb{V}\left(\alpha_N^{-1}\binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i}\nabla_\theta l_{ij}\left(\theta_0\right)\right) &= \alpha_N^{-2}\binom{N}{2}^{-2}\sum_{s=0}^{2}\binom{N}{2}\binom{2}{s}\binom{N-2}{2-s}\Delta_{s,N} \\
&= \frac{\Lambda_0}{n\alpha_N},
\end{aligned}$$

for $\Delta_{s,N}$ the covariance between $\nabla_\theta l_{ij}\left(\theta_0\right)$ and $\nabla_\theta l_{kl}\left(\theta_0\right)$ when $s=0,1,2$ indices are in common. The second equality follows from the fact that $\Delta_{0,N}=\Delta_{1,N}=0$ and the variance calculation above. This suggests that $\hat{\theta}$ converges in mean square to θ_0 at rate $1/n\alpha_N$.

Now, without loss of generality, let $\mathbf{i}=\mathbf{1},\ldots,\mathbf{m}_N$ index those dyads embedded in stable neighborhoods. Here $\{\mathbf{m}_N\}$ is a sequence of integer-valued random variables. Part (i) of Assumption 3 implies that

$$\frac{\mathbf{m}_N}{n\alpha_N}\xrightarrow{p}1,$$

with $n\alpha_N\to\infty$. By Lemma 3 $\{\nabla_\theta l_{\mathbf{i}}\left(\theta_0\right)\}_{\mathbf{i}=\mathbf{1}}^{\mathbf{m}_N}$ are i.i.d. with mean zero and finite variance Λ_0 (from the calculation above). A CLT for a random number of summands (e.g., Theorem 1.9.4 of Serfling (1980)) and Slutsky's Theorem therefore give

$$\frac{1}{\sqrt{n\alpha_N}}\sum_{i=1}^{N}\sum_{j<i}\nabla_\theta l_{ij}\left(\theta_0\right)\xrightarrow{D}\mathcal{N}\left(0,\Lambda_0\right),$$

from which the result follows after another application of Slutsky's Theorem.

B Construction of a stable dyad indicator

It is helpful for computation to have an indicator for neighborhood stability. To construct one first define the matrix

$$\mathbf{S} = \mathbf{D}_1 \circ \mathbf{D}_2 \circ \mathbf{D}_3 + (\iota\iota' - \mathbf{D}_1) \circ (\iota\iota' - \mathbf{D}_2) \circ (\iota\iota' - \mathbf{D}_3) \, .$$

The ij^{th} element of $\mathbf{S}$ equals 1 if the link status of dyad ij is the same across periods 1, 2 and 3 (i.e., either always linked or never linked) and zero otherwise. The diagonal elements of $\mathbf{S}$ equal one by construction. Note that S_{ij}, the ij^{th} element of $\mathbf{S}$, does not correspond to S_{ij} as defined in the main text and Appendix A.

To find *dyads* that (i) change link status across periods 1 and 2, but (ii) leave any remaining links they may have unchanged in periods 1, 2 and 3 define the boolean matrix $\mathbf{T}$ with ij^{th} entry

$$T_{ij} = \begin{cases} 1 & \text{if } D_{1ij} \neq D_{2ij} \;\&\; \sum_{j=1}^{N} S_{ij} = N-1 \;\&\; \sum_{i=1}^{N} S_{ij} = N-1 \\ 0 & \text{otherwise} \end{cases} .$$

To identify agents who do not form or sever *any* links during periods 1 and 2 define the boolean vector $\mathbf{V}$ with i^{th} element

$$V_i = \begin{cases} 1 & \text{if } \sum_{j=1}^{N} D_{1ij}D_{2ij} + (1 - D_{1ij})(1 - D_{2ij}) = N \\ 0 & \text{otherwise} \end{cases} .$$

Use $\mathbf{T}$ and $\mathbf{V}$ to construct an indicator for dyad stability:

$$Z_{ij} = T_{ij} \times \left[\prod_{k \neq i,j} \left(D_{1ik}V_k + (1 - D_{1ik}) \right) \right] \times \left[\prod_{k \neq i,j} \left(D_{1kj}V_k + (1 - D_{1kj}) \right) \right] . \tag{30}$$

Note that $\prod_{k \neq i,j} (D_{1ik}V_k + (1 - D_{1ik})) = 1$ if for all $k \neq i, j$ one of the following two conditions hold: (i) agent i is linked to agent k *and* all of k's links are stable across periods 1 and 2 or (ii) agent i is not linked to k. Hence T_{ij} equals one if the first condition for neighborhood stability applies *and* i and j revise their link status across periods 1 and 2, while the two terms in $[\bullet]$ equal one if the second condition for neighborhood stability holds holds.

References

Ahn, H. and Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1-2):3 – 29.

Almquist, Z. W. and Butts, C. T. (2013). Dynamic network logistic regression: a logistic choice analysis of inter- and intra-group blog citation dynamics in the 2004 us presidential election. *Political Analysis*, 21(4):430 – 448.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Blitzstein, J. and Diaconis, P. (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4):489 – 522.

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.

Chandrasekhar, A. (2016). *The Oxford Handbook of the Economics of Networks*, chapter Econometrics of network formation, pages 303 – 357. Oxford University Press.

Chandrasekhar, A. G. and Jackson, M. O. (2015). A network formation model based on subgraphs. Technical report, Stanford University.

Charlier, E., Melenberg, B., and van Soest, A. H. O. (1995). A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation. *Statistica Neerlandica*, 49(3):324 – 342.

Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Annals of Statistics*, 41(5):2428 – 2461.

Choi, T. Y. and Wu, Z. (2009). Triads in supply networks: theorizing buyer-supplier-supplier relationships. *Journal of Supply Chain Management*, 45(1):8 – 25.

Christakis, N. A., Fowler, J. H., Imbens, G. W., and Kalyanaraman, K. (2010). An empirical model for strategic network formation. Technical Report 16039, National Bureau of Economic Research.

Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society B*, 20(2):215 – 241.

Davis, J. A., Holland, P., and Leinhardt, S. (1971). Comments on professor mazur's hypothesis about interpersonal sentiments. *American Sociological Review*, 36(2):309 – 311.

de Paula, Á. (2016). Econometrics of network models. Technical Report CWP06/16, CEMMAP.

de Paula, Á., Richards-Shubik, S., and Tamer, E. (2015). Identification of preferences in network formation games. Technical Report CWP29/15, CEMMAP.

Dzemski, A. (2014). An empirical model of dyadic link formation in a network with unobserved heterogeneity". Technical report, University of Gothenburg.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129 – 233.

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.

Goodreau, S. M., Kitts, J. A., and Morris (2009). Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks,. *Demography*, 46(1):103 – 125.

Graham, B. S. (2011). *Handbook of Social Economics*, volume 1B, chapter Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers,, pages 965 – 1052. North-Holland.

Graham, B. S. (2013). Comment on "social networks and the identification of peer effects" by paul goldsmith-pinkham and guido w. imbens. *Journal of Business and Economic Statistics*, 31(3):266 – 270.

Graham, B. S. (2014). An econometric model of link formation with degree heterogeneity. Technical Report 20341, National Bureau of Economic Research.

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7:465 – 485.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1360 – 1380.

Gulati, R. and Gargiulo, M. (1999). Where do interorganizational networks come from? *American Journal of Sociology*, 104(5):1439 – 1493.

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'inséé*, 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.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3):293 – 325.

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.

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 –18 97.

Jackson, M. O., Rogers, B. W., and Zenou, Y. (2016). The economic consequences of social network structure. Technical report, Stanford University.

Jackson, M. O. and Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, 106(2):265 – 295.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44 – 74.

Kossinets, G. and Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405 – 450.

Kostin, A. (2010). Probability distribution of distance between pairs of nearest stations in wireless network. *IEEE Electronic 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(3):204 – 213.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3):205 – 228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27(3):313 – 333.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55(2):357 – 362.

Maoz, Z. (2012). Preferential attachment, homophily, and th structure of international networks, 1816 - 2003. *Conflict Management and Peace Science*, 29(3):341 – 369.

Matous, P. and Todo, Y. (2016). Energy and resilience: the effects of endogenous interdependencies on trade network formation across space among major japanese firms. *Network Science*.

McFadden, D. (1974). *Frontiers in Econometrics*, chapter Conditional logit analysis of qualitative choice behavior, pages 105 – 142. Academic Press.

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*, 27:415 – 444.

Mele, A. (2015). A structural model of segregation in social networks. Technical report, John Hopkins University.

Menzel, K. (2015). Strategic network formation with many agents. Technical report, New York University.

Penrose, M. (2003). *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons.

Sheng, S. (2014). A structural econometric analysis of network formation games. Technical report, UCLA.

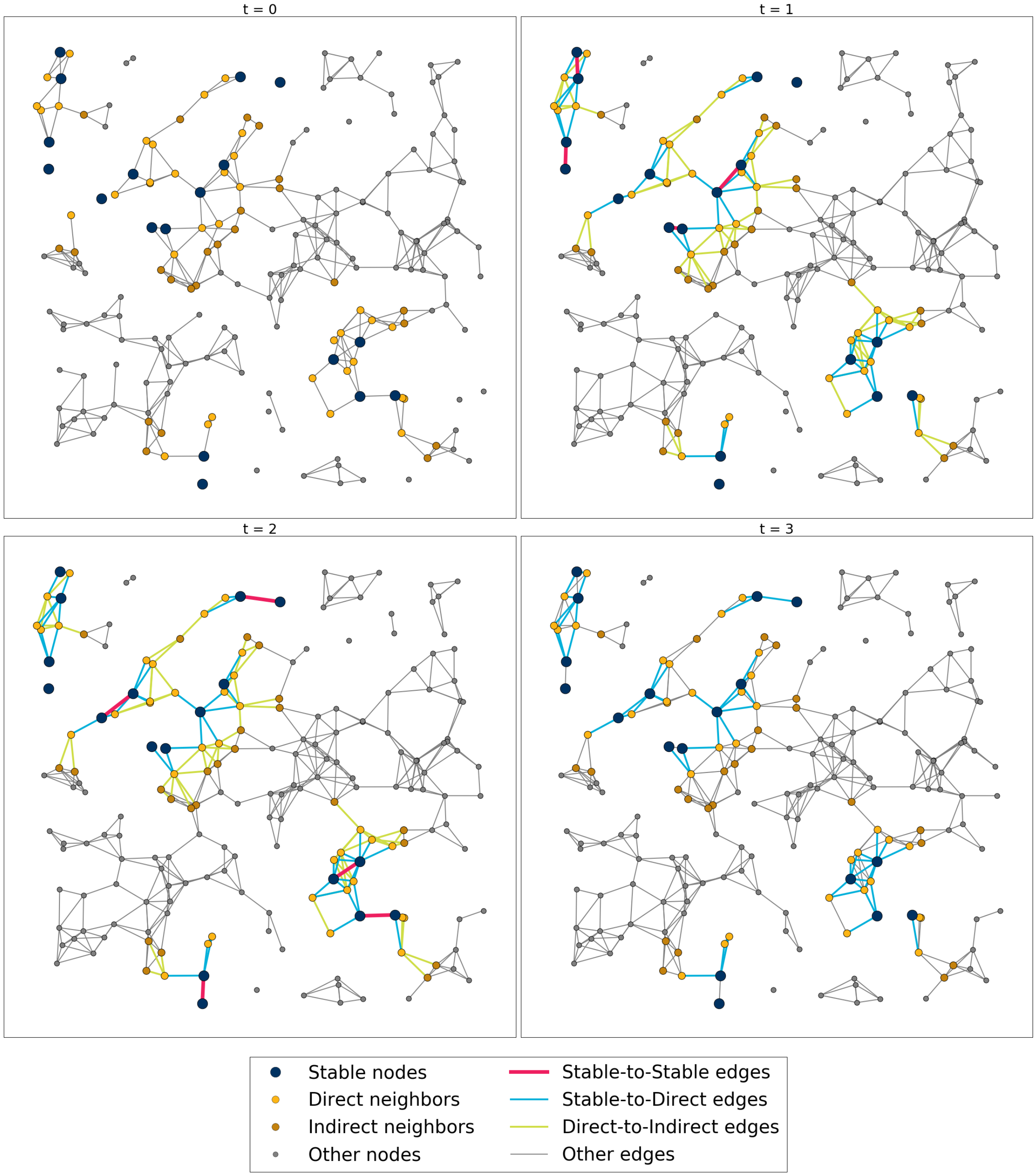
Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, 37:131 – 153.

Snijders, T. A. B. (2013). Network dynamics. In Rafael Wittek, T. S. . V. N., editor, *The Handbook of Rational Choice Social Research*, pages 252 – 279. Stanford University Press.

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.

Watts, A. (2001). A dynamic model of network formation. *Games and Economic Behavior*, 34(2):331 – 341.

Figure 6: Illustrative network sequence for Monte Carlo Design



Notes: A typical sequence of networks for the case where $N = 200$ and average degree in the initial network is 4.