cemmap

centre for microdata methods and practice

An econometric model of network formation with degree heterogeneity

Bryan S. Graham

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP08/17

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

An econometric model of network formation with degree heterogeneity

Bryan S. Graham*

Initial Draft: March 2014, This Draft: December 2016

Abstract

I introduce a model of undirected dyadic link formation which allows for assortative matching on observed agent characteristics (homophily) as well as unrestricted agent level heterogeneity in link surplus (degree heterogeneity). Like in fixed effects panel data analyses, the joint distribution of observed and unobserved agent-level characteristics is left unrestricted. Two estimators for the (common) homophily parameter, β_0, are developed and their properties studied under an asymptotic sequence involving a single network growing large. The first, *tetrad logit* (TL), estimator conditions on a sufficient statistic for the degree heterogeneity. The second, *joint maximum likelihood* (JML), estimator treats the degree heterogeneity $\{A_{i0}\}_{i=1}^{N}$ as additional (incidental) parameters to be estimated. The TL estimate is consistent under both sparse and dense graph sequences, whereas consistency of the JML estimate is shown only under dense graph sequences.

JEL Codes: C31, C33, C35

Keywords: Network formation, homophily, degree heterogeneity, scale-free networks, incidental parameters, asymptotic bias, fixed effects, conditional likelihood, dependent U-Process

*Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880 and National Bureau of Economic Research, e-mail: bgraham@econ.berkeley.edu, web: `http://bryangraham.github.io/econometrics/`. The first draft of this paper dates to March 2014 and was titled "An empirical model of network formation: detecting homophily when agents are heterogenous". A Python 2.7.11 module implementing the methods described in this paper is available for download from my web page. I thank Jinyong Hahn for sharing an unpublished appendix of Hahn and Newey (2004), Joachim De Weerdt for generously making his Nyakatoke dataset available, Eric Auerbach, Peter Bickel, Stephane Bonhomme, Jesus Carro, Gary Chamberlain, Xiaohong Chen, Nicholas Christakis, Ivan Fernandez-Val, Jim Heckman, Yuichi Kitamura, Elena Manresa, Costas Meghir, Roger Moon, Jong-Myun, Matt Shum, Martin Weidner and Bin Yu for numerous helpful suggestions. This paper has also benefited from the reactions of audiences at the 2014 Seattle-Vancouver Econometrics Conference, the 2014 All California Econometrics Conference, the 2015 Econometrics Journal Conference on Networks, the 2016 BeNet Conference at UCL/CORE, as well as from participants in seminars and workshops at New York University, UC Berkeley, the Toulouse School of Economics, Oxford, Cambridge, St. Gallen, CEMFI, the Yale Institute of Network Sciences, Stanford, USC, UCSD, Chicago, Michigan, Brown and Harvard-MIT. This revision incorporates many useful suggestions made by the co-editor, Elie Tamer, and four anonymous referees. Finally, I am especially grateful for helpful conversations and suggestions from my three immediate colleagues in the Economics Department at Berkeley: Michael Jansson, Demian Pouzo and Jim Powell. Financial support from NSF grant SES #1357499 is gratefully acknowledged. All the usual disclaimers apply.

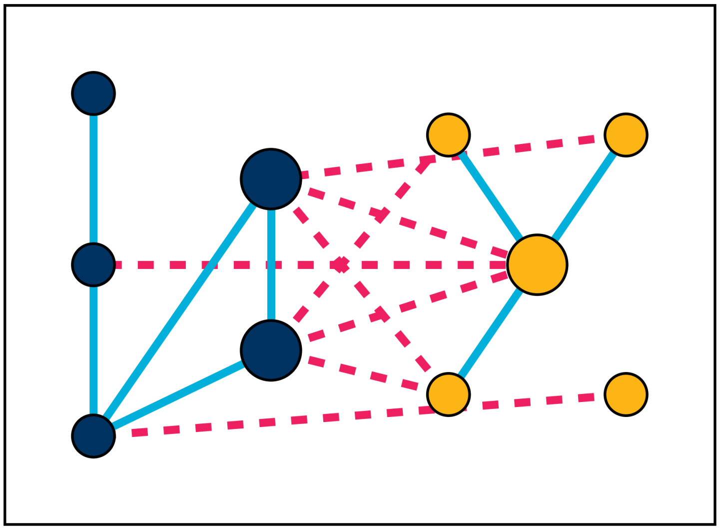
Homophily, the tendency of individuals to form connections with those like themselves, is a widely-observed feature of real world social and economic networks (e.g., McPherson et al., 2001). Equally common is degree heterogeneity: variation in the number of links (i.e., degree) across individuals. In particular, the conjunction of many low degree individuals with few links, and a handful of high degree "hub" individuals with many links, characterizes many networks (e.g., Barabási and Albert, 1999). The presence and magnitude of homophily and degree heterogeneity has implications for how information diffuses, the spread of epidemics, as well as the speed and precision of social learning (e.g., Pastor-Satorras and Vespignani, 2001; Jackson and Rogers, 2007; Golub and Jackson, 2012; Jackson and López-Pintado, 2013).[1]

This paper formulates and studies a model of link formation that flexibly accommodates both homophily and degree heterogeneity. To motivate the model, as well as heuristically introduce some of the identification issues involved, consider the small network depicted in Figure 1. This network consists of two observable types of agents, colored "Berkeley Blue" (dark) and "California Gold" (light) in the figure. There are also three "hub" agents in the network (the larger nodes in the center of the graph). Whether an agent is a hub or not is unobserved by the researcher. This network is a random draw from a population characterized by a strong structural taste for homophily (see Section 1 below and the notes to the figure for details). Although homophily underlies the network formation process, half of all links – specifically eight out of sixteen – are heterophilic (i.e., between agents of a different type/color). Since the frequency of links across (homphilous) dyads of the same type is identical to that across (heterophilous) dyads of differing types a researcher fitting common models of link formation to these data might fail to conclude that preferences are, in fact, homophilic. The presence of hub agents, who form many links irrespective of type, effectively attenuates measured homophily.

The model outlined below is designed to help researchers avoid this type of inferential mistake. It augments a standard dyadic model of link formation, as used by, for example, Fafchamps and Gubert (2007), Lai and Reiter (2000), Apicella et al. (2012) and Attanasio et al. (2012), with agent-specific unobserved degree heterogeneity. Specifically agents freely vary in the generic surplus they generate when forming a match. The surplus associated with any given match may further vary with observable characteristics of the dyad. For example surplus may be systematically higher between agents who are close in age (homophily on age). Unlike prior work incorporating degree heterogeneity (e.g., van Duijn et al., 2004; Krivitsky et al., 2009), the joint distribution of the unobserved degree heterogeneity and

[1] Apicella et al. (2012) even study the relationship between homophily and the emergence of cooperation in hunter-gatherer societies.

Figure 1: Homophily and degree heterogeneity



Notes: See Section 1 below for additional details on the construction of the figure as well as notational definitions. The figure shows a simulated network with both correlated degree heterogeneity and homophilous link formation. California Gold (light) and Berkeley Blue (dark) shaded nodes respectively denote $X = -1$ and $X = 1$ agents. Smaller nodes denote "low degree agents" (with $A = \underline{a}$) and larger nodes "high degree" or "hub" agents (with $A = \bar{a}$). Heterophilic and homophilic links are depicted with, respectively, dashed and solid lines. Links form according to equation (1) of Section 1 with $\Pr(X = -1, A = \underline{a}) = 0.4$, $\Pr(X = -1, A = \bar{a}) = 0.1$, $\Pr(X = 1, A = \underline{a}) = 0.3$ and $\Pr(X = 1, A = \bar{a}) = 0.2$, $\beta_0 = 1$, $W_{ij} = X_i X_j$ and $\underline{a}$ and $\bar{a}$ chosen such that a $(X = -1, A = \underline{a})$ to $(X = 1, A = \underline{a})$ link occurs with probability 0.025 and a $(X = -1, A = \bar{a})$ to $(X = 1, A = \bar{a})$ link occurs with probability 0.8.

observed agent attributes is left unrestricted. The treatment here is a "fixed effects" one (Chamberlain, 1980, 1985). This allows for settings similar to that depicted in Figure 1, where dark "Berkeley Blue" agents are more likely to be hubs than light "California Gold" ones.[2]

In the model each agent $i = 1, \ldots, N$ has an individual-specific "degree effect", A_i. If these effects are treated as (incidental) parameters, then the dimension of the parameter vector grows with the number of agents in the network. This makes the estimation problem non-standard. Textbook results on the large sample properties of maximum likelihood estimates (MLEs) do not apply (e.g., Neyman and Scott, 1948). In this paper I introduce and study two fixed effects estimators of the common parameters characterizing homophily. The first estimator implicitly conditions on a sufficient statistic for the degree effects. The second estimates the degree effects jointly with the common parameters.

[2]De Weerdt (2004, Column 3, Table 7) fits the JML estimator described below to a risk sharing network from Tanzania. He does not analyze the asymptotic sampling properties of the JMLE.

The first estimator is based on a standard application of minimal sufficiency in exponential families (Andersen, 1973). Similar results form the basis of conditional maximum likelihood estimators in nonlinear panel data models (Cox, 1958; Chamberlain, 1980). Recently, in independent work, Charbonneau (2014) uses sufficiency arguments to develop conditional estimators for nonlinear models with multiple fixed effects. Her analysis is inspired by empirical studies of international trade, where the introduction of importer and exporter effects is common (e.g., Santos Silva and Tenreyro, 2006). While not explicitly formulated as such, the implicit network structure of her model is one with directed edges ("does country i export to country j?"). In contrast, the results presented here apply to undirected networks. Charbonneau (2014) does not characterize the large sample properties of her estimator.[3]

The conditional estimator I introduce below is based on the relative frequency of different types of subgraphs, each consisting of four agents (called tetrads). I call this estimator the *tetrad logit* (TL) estimator. The tetrad logit criterion function is related to the class of U-Process minimizers studied by Honoré and Powell (1994) among others. Unfortunately, these prior results do not apply to the TL estimator. Although the TL criterion function consists of a sum over all $\binom{N}{4}$ quadruples of agents, analogous to a 4th order U-process, its kernel includes random variables defined at both the agent *and* dyad level. The criterion function is therefore not a U-statistic, conventionally defined. Nevertheless I adapt various tools from the literature on U-Statistics to characterize the asymptotic properties of the TL estimator. Specifically, the tetrad logit "score" vector behaves similarly to a fourth order degenerate U-statistic. A Hoeffding (1948) variance decomposition indicates that this degeneracy is of order one such that the leading variance term is inversely proportional to the number of dyads in the network, $n \stackrel{def}{=} \binom{N}{2} = \frac{1}{2}N(N-1)$, not the number of agents, N.[4] This score vector is asymptotically equivalent to a certain projection which involves summation over dyads (not agents). This projection, however, is not a sum of independent components. Fortunately its summands have a particular conditional independence structure, which I exploit to demonstrate asymptotic normality by adapting a construction due to Chatterjee (2006).

The second estimator jointly estimates the common and incidental parameters by maximum likelihood. I call this estimator the *joint maximum likelihood* (JML) estimator. The key insight is that, although the number of parameters is of order N, the number of conditionally independent log-likelihood components is of order N^2. Each dyad contributes for a total of $n = \frac{1}{2}N(N-1) = O(N^2)$ log-likelihood components. Since the amount of "data" is

[3]In a recent paper, which builds, in part, on the results presented below, Jochmans (2016a) formalizes the approach of Charbonneau (2014) and provides large sample results.

[4]This statement only applies exactly to dense network sequences, the sparse case is more complicated, as detailed below.

increasing at a rate faster than the dimension of the parameter, the joint maximum likelihood estimates of the common parameters are consistent, however, their limit distribution is biased. Accurate inference therefore requires bias-correction. This analysis parallels recent findings from the non-linear panel data literature under large-N, large-T asymptotics (e.g., Hahn and Newey, 2004; Arellano and Hahn, 2007). Dzemski (2014), in related work, studies the properties of joint maximum-likelihood applied to the directed network model of Charbonneau (2014). His analysis builds on Fernández-Val and Weidner's (2016) study of non-linear panel data models with both individual- and time-effects. In contrast, the technical details of the analysis presented here draws from Chatterjee et al.'s (2011) analysis of the β-model of network formation (cf., Yan and Xu, 2013)[5] and Hahn and Newey (2004).

I demonstrate consistency and asymptotic normality of the TL and JML estimators under differing regularity conditions. In both cases results are established under an asymptotic sequence involving a single network which grows in size. To my knowledge, the two estimators introduced here represent the first frequentist analyses of an econometric model of link formation under "single network asymptotics".[6] The TL estimate is shown to be consistent under both sparse graph sequences, where the number of links per agent is bounded, as well as dense graph sequences, where the number of links per agent is proportional to the total number of agents in the limit. The JML estimate is only shown to be consistent under dense graph sequences. This difference is likely to be consequential in ways relevant to empirical researchers. Many social and economic networks are "sparse", in the sense that only a small fraction of all possible links are present, the JML estimator may have poor finite sample properties in such settings (a conjecture I explore through a series Monte Carlo experiments summarized in the Supplemental Materials). An advantage of the JML estimator, relative to the TL one, is that it produces estimates of the incidental as well as the common parameters. This allows for computation of marginal effects and counterfactuals. The two estimators are complementary, with the TL estimator being applicable to a wider class of problems, but the JML estimator providing consistent estimates of more features of the network generating process.

[5]This work is, in turn, closely connected to an older literature on the Bradley-Terry model of paired comparisons (e.g., Simons and Yao, 1998, 1999).

[6]Prior empirical work based on a single network has generally taken a Bayesian approach (e.g., van Duijn et al., 2004; Krivitsky et al., 2009; Christakis et al., 2010; Mele, 2013; Goldsmith-Pinkham and Imbens, 2013). Extant frequentists analyses involve asymptotic sequences based upon an increasing number of independent networks (e.g., Miyauchi, 2016; Sheng, 2014). Chandrasekhar and Jackson (2015) do work under single network asymptotics, but in the context of a rather different model from the one considered here. Leung (2015) also develops some tools for frequentist inference based on a single large network. Since the initial draft of this paper was circulated, several additional working papers have appeared building on various results presented below, as well as prior work by others; these papers include those by Dzemski (2014); Candelaria (2016); Jochmans (2016b,a); Nadler (2015) and Yan et al. (2016).

An important limitation of the analysis presented here is that it rules out interdependent link preferences, whereby agents' preferences over a link may vary with the presence or absence of links elsewhere in the network. The study of network formation in the presence of interdependent preferences is one theme of recent theoretical research on networks (e.g., Jackson and Wolinsky, 1996; Bala and Goyal, 2000; Jackson and Watts, 2002). Christakis et al. (2010); Mele (2013); Goldsmith-Pinkham and Imbens (2013); Graham (2013, 2016); Menzel (2016) and de Paula et al. (2015) are some recent attempts to study econometric models of network formation with interdependent preferences under various assumptions. None of these papers, with the exception of Goldsmith-Pinkham and Imbens (2013) and Graham (2013, 2016), incorporate correlated unobserved agent heterogeneity into their modeling frameworks, as is done here. In Section 3 I discuss how to extend the results presented below to incorporate interdependent preferences in link formation (at least of a certain limited type) when the network is observed for two or more periods.

Section 1 formally introduces a dyadic model of link formation with degree heterogeneity and presents a baseline set of maintained assumptions. Section 2 presents the tetrad logit (TL) and joint maximum likelihood (JML) estimators and characterizes their large sample properties. Section 3 sketches several extensions of the basic model. Proofs associated with the tetrad logit estimate can be found in Appendix A. All remaining proofs, as well as a Monte Carlo analysis of the two estimators finite sample properties, can be found in the Supplemental Materials.

Notation

In what follows 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. If $\mathbf{B}$ is an $N \times N$ matrix with $(i,j)^{th}$ element b_{ij}, then $\|\mathbf{B}\|_{\max} = \sup_{i,j} |b_{ij}|$, $\|\mathbf{B}\|_{\infty} = \sup_i \sum_{j=1}^{N} |b_{ij}|$, and $\|\mathbf{B}\|_{2,1} = \sum_{j=1}^{n} \left[\sum_{i=1}^{m} |b_{ij}|^2\right]^{1/2}$. I use ι_N to denote a $N \times 1$ vector of ones and I_N the $N \times N$ identity matrix. The notation $\sum_{i<j<k}$ is a shorthand for $\sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N}$. A "0" subscript on a parameter denotes its population value and may be omitted when doing so causes no confusion.

1 Model and baseline assumptions

Consider a large population of potentially connected agents. Depending on the context agents may be individuals, households, firms, or nation-states (among many other types of possible actors). Let $i = 1, \ldots, N$ index a random sample of size N from this population. Each of the $n \stackrel{\text{def}}{\equiv} \binom{N}{2} = \frac{1}{2}N(N-1)$ pairs of sampled agents constitute a dyad. For each ij dyad let $D_{ij} = 1$ if i and j are connected and zero otherwise.[7] Connections are undirected (i.e., $D_{ij} = D_{ji}$) and self-ties are ruled out (i.e., $D_{ii} = 0$). The $N \times N$ matrix $\mathbf{D}$, with ij^{th} element D_{ij}, is called the adjacency matrix. This matrix is binary and symmetric, with zeros on its main diagonal. The adjacency matrix encodes the structure of links across all sampled agents. It what follows I will refer to a set of such links as, equivalently, a network or graph.

An agent's *degree* equals the number of links she has: $D_{i+} = \sum_{j \neq i} D_{ij}$ (the "+" denotes "leave-own-out" summation over the replaced index). The row (or column) sums of the adjacency matrix, denoted by the $N \times 1$ vector $\mathbf{D}_+ = (D_{1+}, \ldots, D_{N+})'$, give the network's *degree sequence.*

The econometrician also observes X_i, a vector of agent-level attributes. These agent-level attributes are used to construct the $K \times 1$ dyad-level vector $W_{ij} = g(X_i, X_j)$. The function $g(\cdot, \cdot)$ is symmetric in its arguments so that $W_{ij} = W_{ji}$. As an example if X_{1i} and X_{2i} are location coordinates, then $W_{ij} = \left((X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2\right)^{1/2}$ equals the distance between i and j.

Agents i and j form a link if the total surplus from doing so is positive:

$$D_{ij} = \mathbf{1}\left(W_{ij}'\beta_0 + A_i + A_j - U_{ij} \geq 0\right), \tag{1}$$

where $\mathbf{1}(\bullet)$ denotes the indicator function. Link surplus consists of three components:

1. a systematic component which varies with observed dyad attributes, $W_{ij}'\beta_0$ (*homophily*),
2. a component which varies with the unobserved agent-level attributes $\{A_i\}_{i=1}^N$, (*degree heterogeneity*) and
3. an idiosyncratic component, $U_{ij} = U_{ji}$, assumed independently and identically distributed across dyads.

Because links are undirected, the surplus function is specified to ensure that the linking rule for D_{ij} coincides with that for D_{ji}. This requires, as noted above, that $W_{ij} = W_{ji}$, but also that A_i and A_j enter (1) symmetrically. Finally, observe that any components of surplus linear in X_i and X_j will be absorbed by the degree heterogeneity terms $\{A_i\}_{i=1}^N$.

Implicit in rule (1) is the presumption that utility is transferable across *directly linked* agents; all links with positive net surplus form (Bloch and Jackson, 2007). Rule (1) re-

[7] Connections may be equivalently referred to as links, ties, friendships, or edges depending on the context.

sults in a complete and coherent model of network formation. For a given draw of $\mathbf{U} = (U_{12}, U_{13}, \ldots, U_{N-1N})'$ the network is uniquely determined.

Baseline assumptions

Let $\mathbf{X}$ be the $N \times \dim(X)$ matrix of observed agent attributes and $\mathbf{A}_0$ the $N \times 1$ vector of unobserved agent-level degree heterogeneity terms. All of the results presented below maintain the following three assumptions, with additional assumptions made for specific results.

Assumption 1. (Likelihood) The conditional likelihood of the network $\mathbf{D} = \mathbf{d}$ is

$$\Pr(\mathbf{D} = \mathbf{d} | \mathbf{X}, \mathbf{A}_0) = \prod_{i<j} \Pr(D_{ij} = d | X_i, X_j, A_{i0}, A_{j0})$$

with

$$\Pr(D_{ij} = d | \mathbf{X}, \mathbf{A}_0) = \left[\frac{1}{1 + \exp\left(W_{ij}'\beta_0 + A_{i0} + A_{j0}\right)}\right]^{1-d} \left[\frac{\exp\left(W_{ij}'\beta_0 + A_{i0} + A_{j0}\right)}{1 + \exp\left(W_{ij}'\beta_0 + A_{i0} + A_{j0}\right)}\right]^{d}$$

for all $i \neq j$.

Assumption 1 implies that the idiosyncratic component of link surplus, U_{ij}, is a standard logistic random variable that is independently and identically distributed across dyads. The logistic assumption is important for the tetrad logit (TL) estimator, but less so for the joint maximum likelihood (JML) estimator (although my proof strategy does make use of the logit structure extensively in both cases).

Assumption 1 also implies that links form independently conditional on $\mathbf{X}$ and $\mathbf{A}_0$. Consider the agents i, j and k. Conditional on these agents' observed and unobserved characteristics, respectively X_i, X_j, X_k and A_i, A_j, A_k, the events "i and j are connected", "i and k are connected" and "j and k are connected" are independent of one another.

Importantly independence is conditional on the *latent* agent attributes $\{A_i\}_{i=1}^{N}$. Unconditionally on these attributes, independence does not hold. For example, conditioning on X_i, X_j, X_k, but not on A_i, A_j, A_k, observing that "i and j are connected" increases the ex ante probability placed on the event "i and k are connected". Dependence of this type is generated by the presence of A_i in both the ij and ik linking rules. This type of dependence is analogous to that allowed for by a strict exogeneity assumption in a single agent static panel data model (e.g., Chamberlain, 1984).

The assumption that links form independently of one another conditional on agent attributes will be plausible in some settings, but not in others. Specifically, rule (1) and Assumption 1 are appropriate for settings where the drivers of link formation are predominately bilateral in nature, as may be true in some types of friendship and trade networks as well as in models of (some types of) conflict between nation-states (e.g., Santos Silva and Tenreyro, 2006; Fafchamps and Gubert, 2007; Lai and Reiter, 2000). In such settings, as outlined below, the inclusion of unobserved agent attributes in the link formation model is a significant, and useful, generalization relative to many commonly-used models.

In other settings, however, link decisions may have strong strategic aspects. For example, Apple may prefer that its supply-chain not overlap with Samsung's (in order to protect manufacturing know-how). In such settings the events "firm A supplies Samsung" and "firm A supplies Apple" will not be independent. With strategic interaction, the presence or absence of a link in one part of the network may structurally influence the returns to link formation in other parts of the network. Such interdependencies generate interesting challenges that are not addressed here. The surveys by Graham (2015) and de Paula (2016) provide additional discussion as well as references.

The approach taken here is to study identification and estimation issues when links form according to rule (1) and Assumption 1. This setting both covers a useful class of empirical examples, and represents a natural starting point for formal econometric analysis. An analogy with single agent discrete choice panel data models is perhaps useful. In that setting early methodological work focused on introducing unobserved correlated heterogeneity into static models of choice (e.g., Chamberlain, 1980; Manski, 1987). Later work subsequently incorporated a role for state dependence in choice (e.g., Chamberlain, 1985; Honoré and Kyriazidou, 2000). Section 3 sketches some extensions of the framework developed here to incorporate certain types of interdependencies in link formation.

Assumption 2. (Compact Support)
(i) $\beta_0 \in \text{int}(\mathbb{B})$, with $\mathbb{B}$ a compact subset of $\mathbb{R}^K$.
(ii) the support of W_{ij} is $\mathbb{W}$, a compact subset of $\mathbb{R}^K$.

Part (i) of Assumption 2 is standard in the context of nonlinear estimation problems. Together with part (ii) it implies that the observed component of link surplus, $W_{ij}'\beta_0$ will have bounded support. This simplifies the proofs of the main Theorems, especially those of the JML estimator, as will be explained below. For the tetrad logit estimator part (ii) could be relaxed by assuming, instead, that W_{ij} has a sufficient number of bounded moments.

Assumption 3. (Random Sampling) Let $i = 1, \ldots, N$ index a random sample of agents from a population satisfying Assumptions 1 and 2. The econometrician observes (D_{ij}, W_{ij}) for $i = 1, \ldots, N$, $j < i$ (i.e., for all sampled dyads).

Network data can be difficult and expensive to collect, consequently many analyses in the social sciences are based on incomplete graphs (e.g., Banerjee et al., 2013). One implication of Assumption 3 is that estimation and inference may be based upon only a subset of the full network.[8]

2 Estimation

Tetrad logit (TL) estimation

The tetrad logit estimator is based on an identifying implication of the model defined by (1) and Assumptions 1 through 3 that is invariant to $\{A_{i0}\}_{i=1}^{N}$. To derive this implication rewrite the conditional likelihood of the event $\mathbf{D} = \mathbf{d}$ given $(\mathbf{X}, \mathbf{A}_0)$, as

$$\Pr(\mathbf{D} = \mathbf{d} | \mathbf{X}, \mathbf{A}_0) = \prod_{i<j} \left[\frac{\exp\left(W_{ij}'\beta_0 + T_{ij}'\mathbf{A}_0\right)}{1 + \exp\left(W_{ij}'\beta_0 + T_{ij}'\mathbf{A}_0\right)} \right]^{d_{ij}} \left[\frac{1}{1 + \exp\left(W_{ij}'\beta_0 + T_{ij}'\mathbf{A}_0\right)} \right]^{1-d_{ij}},$$

where T_{ij} is an $N \times 1$ vector with a one for its i^{th} and j^{th} elements and zeros elsewhere such that $T_{ij}'\mathbf{A}_0 = A_{0i} + A_{0j}$. After some manipulation this likelihood can be put into the exponential family form

$$\Pr(\mathbf{D} = \mathbf{d} | \mathbf{X}, \mathbf{A}_0) = c(\mathbf{X}; \beta_0, \mathbf{A}_0) \exp\left(S_1(\mathbf{d}, \mathbf{X})' \beta_0\right) \exp\left(S_2(\mathbf{d})' \mathbf{A}_0\right) \tag{2}$$

where

$$S_1(\mathbf{d}, \mathbf{X}) = \sum_{i=1}^{N} \sum_{j<i} d_{ij} W_{ij},\ S_2(\mathbf{d}) = \begin{pmatrix} d_{1+} & \cdots & d_{N+} \end{pmatrix}'.$$

Inspection of (2) indicates that $\mathbf{D}_+ = (D_{1+}, \ldots, D_{N+})'$, the network's degree sequence, is a sufficient statistic for $\mathbf{A}_0$.

An important strand of network research takes the degree sequence as its primary object of interest, since many other topological features of networks are fundamentally constrained by it (e.g., Albert and Barabási, 2002).[9] For example, Graham (2015) shows that the mean and

[8] Shalizi and Rinaldo (2013) call this property "consistency under sampling".

[9] Faust (2007) develops this point empirically using a large database of social networks. Newman (2010) refers to the degree distribution as one of the "...most fundamental of network properties..." (p. 243).

variance of a network's degree sequence can be expressed as a function of its triad census (i.e., the number of triads with no links, one link, two links and three links).[10]

Let $\mathbb{D}^s$ denote the set of all feasible network adjacency matrices with degree sequence $\mathbf{D}_+ = \mathbf{d}_+$:

$$\mathbb{D}^s = \left\{\mathbf{v} \,:\, \mathbf{v} \in \mathbb{D},\, S_2\left(\mathbf{v}\right) = S_2\left(\mathbf{d}\right)\right\}.$$

Here $\mathbb{D}$ denotes the set of all $2^{\binom{N}{2}}$ undirected binary adjacency matrices. Solving for the conditional probability of the observed network given its degree sequence yields

$$\Pr\left(\mathbf{D} = \mathbf{d} \middle| \mathbf{X}, \mathbf{A}_0, S_2\left(\mathbf{D}\right) = S_2\left(\mathbf{d}\right)\right) = \frac{\exp\left(\sum_{i=1}^{N}\sum_{j<i} d_{ij}W_{ij}'\beta_0\right)}{\sum_{\mathbf{v}\in\mathbb{D}^s}\exp\left(\sum_{i=1}^{N}\sum_{j<i} v_{ij}W_{ij}'\beta_0\right)}, \qquad (3)$$

which does not depend on $\mathbf{A}_0$.

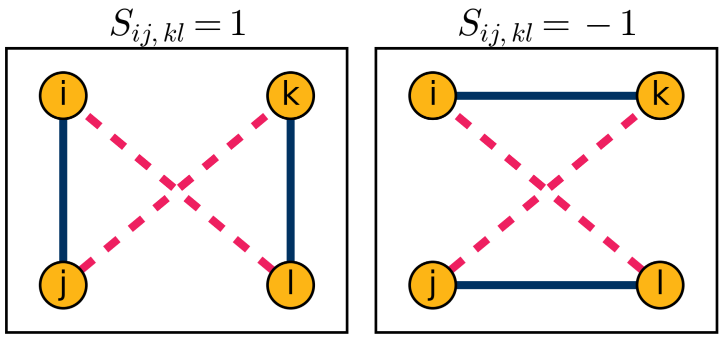
The model defined by (1) and Assumptions 1 to 3 allows for arbitrary degree sequences and hence can replicate many types of network structures. A loose intuition, implicit in the form of the conditional likelihood (3), is that the heterogeneity parameters $\{A_{0i}\}_{i=1}^{N}$ tie down the degree distribution of the network (i.e., how many ones/links are present in each row (or column) of $\mathbf{D}$). The precise location of each link *within a given row/column* is then driven by variation in $W_{ij}'\beta_0$.

Even for small networks, consisting of say a few dozen agents, the set $\mathbb{D}^s$ will typically be far too large to enumerate such that (3) cannot be exactly evaluated. Blitzstein and Diaconis (2011) derive a method for sampling uniformly from $\mathbb{D}^s$, which could be used to estimate (3) via simulation. The analysis of the resulting simulated conditional maximum likelihood estimate of β_0 would be an interesting topic for future research. Here I instead form an estimator based on the relative probability of different types of subgraph configurations. While this approach is unlikely to be as efficient as one based directly on (3), it has the advantage of yielding a criterion function that is easy to evaluate and maximize.

Figure 2 depicts two tetrad configurations. In the first (left) subgraph the ij and kl edges are present, but the ik and jl ones are not. In the second (right) subgraph the opposite configuration is observed. Edges il and jk, depicted as dashed lines in the figure, may or may not be present. However, if they are present, they are assumed to be so in both subgraphs. The two subgraphs, when the dashed edges are omitted, share identical degree sequences of $(1, 1, 1, 1)'$. Because a rewiring from the left-hand subgraph to the right-hand subgraph leaves its degree sequence unchanged, the relative probability of observing one

[10]Holland and Leinhardt (1976) develop this point more comprehensively in the context of directed networks or digraphs.

Figure 2: Edge Swaps and the Definition of $S_{ij,kl}$



Notes: The figure depicts tetrad configurations consistent with the events $S_{ij,kl} = 1$ (left) and $S_{ij,kl} = -1$ (right). Solid lines denote required edges, dashed lines denote edges, the presence or absence of which, do not affect the value of $S_{ij,kl}$. However, if they are present, they are assumed to be so in both subgraphs.

subgraph or the other – *conditional on observing one of them* – will not depend on $\mathbf{A}_0$. The tetrad logit estimator is constructed from this implication.

To be precise, let $S_{ij,kl} = 1$ if we observe the edges ij and kl, but *not* ik and jl, -1 if we observe the opposite, and zero otherwise (see Figure 2). We can construct $S_{ij,kl}$ from the adjacency matrix as

$$S_{ij,kl} = D_{ij}D_{kl}\left(1 - D_{ik}\right)\left(1 - D_{jl}\right) - \left(1 - D_{ij}\right)\left(1 - D_{kl}\right)D_{ik}D_{jl}. \tag{4}$$

Since subgraph configurations with $S_{ij,kl} = 1$ and $S_{ij,kl} = -1$ share the same (subgraph) degree sequence, the conditional probability

$$\Pr\left(S_{ij,kl} = 1 \middle| \mathbf{X}, \mathbf{A}_0, S_{ij,kl} \in \{-1, 1\}\right) = \frac{\exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}{1 + \exp\left(\tilde{W}'_{ij,kl}\beta_0\right)}, \tag{5}$$

with $\tilde{W}_{ij,kl} = W_{ij} + W_{kl} - (W_{ik} + W_{jl})$, does not depend on A_{i0}, A_{j0}, A_{k0} or A_{l0}. The form of (5) accords with the heuristic intuition given above. The contribution of unobserved heterogeneity to total net surplus is the same for the two subgraphs shown in Figure 2, hence the (conditional) frequency with which each is observed depends only on the amount of "observable" surplus associated with each. If $\tilde{W}'_{ij,kl}\beta_0 > 0$, then the observable surplus associated with configuration one ($S_{ij,kl} = 1$) exceeds that associated with configuration two ($S_{ij,kl} = -1$):

$$\left(W_{ij} + W_{kl}\right)'\beta_0 > \left(W_{ik} + W_{jl}\right)'\beta_0,$$

and hence the left-hand configuration in Figure 2 is observed more frequently than the right-hand one.

The index in (5) takes an "increasing difference" form, highlighting the close connection between homophily in matching and structural complementarity in preferences (cf., Fox, 2010; Graham, 2011).

The conditional log-likelihood associated with configuration $S_{ij,kl}$ is

$$l_{ij,kl}\left(\beta_0\right) = \left|S_{ij,kl}\right| \left\{ S_{ij,kl}\tilde{W}'_{ij,kl}\beta_0 - \ln\left[1+\exp\left(S_{ij,kl}\tilde{W}'_{ij,kl}\beta_0\right)\right]\right\}. \tag{6}$$

Object (6) is not invariant to permutations of its indices. To impose symmetry I sum $l_{ij,kl}\left(\beta\right)$ across all possible permutations of its indices, yielding

$$g_{ijkl}\left(\beta\right) = \frac{1}{4!}\sum_{\pi\in\Pi_4} l_{\pi_1\pi_2,\pi_3\pi_4}\left(\beta\right), \tag{7}$$

with Π_4 the group of all $4! = 24$ permutations of a 4 element vector. The kernel $g_{ijkl}(\beta)$ is symmetric in its arguments. Averaging across permutations to induce index symmetry is common in the literature on U-Statistics (e.g., Ferguson, 2005).

The tetrad logit criterion function consists of a summation of contributions (7) over all $\binom{N}{4}$ distinct tetrads in the network. That is $\hat{\beta}_{\mathrm{TL}}$ maximizes

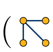
$$L_N\left(\beta\right) = \binom{N}{4}^{-1}\sum_{i<j<k<l} g_{ijkl}\left(\beta\right). \tag{8}$$

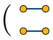
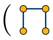
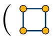
This estimate satisfies the first order condition

$$\nabla_\beta L_N\left(\hat{\beta}_{\mathrm{TL}}\right) = \binom{N}{4}^{-1}\sum_{i<j<k<l} s_{ijkl}\left(\hat{\beta}_{\mathrm{TL}}\right) = 0, \tag{9}$$

where $s_{ijkl}\left(\beta\right) \overset{\mathrm{def}}{=} \nabla_\beta g_{ijkl}\left(\beta\right)$.

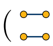
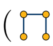
The anatomy of the tetrad logit criterion function

A tetrad can be wired in up to $2^6 = 64$ different ways. Each of these wirings belong to one of the 11 isomorphism classes shown in Figure 3. Forty six of the 64 possible wirings, falling into 8 of the 11 isomorphism classes, are completely determined by their (subgraph) degree sequence. Four example, there are four isomorphisms of the three star configuration () each with a degree sequence equal to (a permutation of) $(3,1,1,1)'$.

The remaining 18 possible wirings *share* their degree sequence with at least one other wiring. These wirings fall into either the two edge (), four path (), or four cycle () isomorphism classes. All three two edge isomorphisms share the degree sequence $(1,1,1,1)'$. There are a total of 12 different four path isomorphisms. For each of the six unique permutations of the degree sequence $(2,2,1,1)'$ there are two feasible wirings, for 12 wirings in total. Finally, there are three isomorphisms of the four cycle configuration, all with a degree sequence of $(2,2,2,2)'$. The tetrad $\{i,j,k,l\}$ only makes a non-zero contribution to the tetrad logit criterion function (8) if it is wired in one of the 18 ways associated with a non-unique degree sequence. These are the only tetrads which can be used to identify β_0, since all other tetrads have no variation in structure conditional on their degree sequence.

Although there are 24 possible permutations of the index set $\{i,j,k,l\}$, it is straightforward, although tedious, to verify that the summand in the tetrad kernel $g_{ijkl}(\beta)$ takes, at most, three different values. This follows since, for example, $l_{ij,kl}(\beta) = l_{ik,jl}(\beta) = l_{ji,lk}(\beta) = l_{jl,ik}(\beta) = l_{kl,ij}(\beta) = l_{ki,lj}(\beta) = l_{lk,ji}(\beta) = l_{lj,ki}(\beta)$. This gives the simplified kernel

$$g_{ijkl}(\beta) = \frac{1}{3}\left[l_{ij,kl}(\beta) + l_{ij,lk}(\beta) + l_{ik,lj}(\beta)\right] \tag{10}$$

for $l_{ij,kl}(\beta)$ as defined in (6). If $\{i,j,k,l\}$ belongs to either the two edge (), four path (), or four cycle () isomorphism class, then at least one component of (10) will be non-zero and hence the tetrad will make a non-trivial contribution to the tetrad logit estimate.

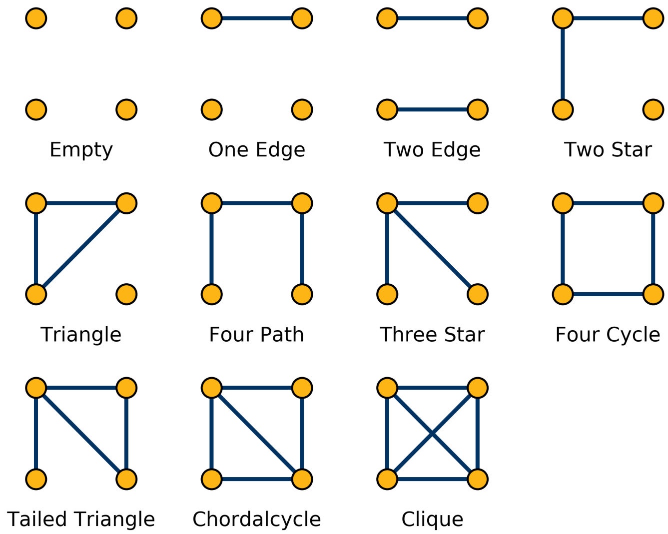
Large sample properties of $\hat{\beta}_{\mathrm{TL}}$

As asserted in the introduction, one attractive feature of $\hat{\beta}_{\mathrm{TL}}$ is that it remains consistent and asymptotically normal under sequences of networks that are sparse in the limit. Developing this claim requires additional notation. For a network of size N, the marginal probability of a randomly sampled dyad linking is given by

$$\rho_N = \Pr(D_{ij} = 1) = \mathbb{E}\left[\Pr(D_{ij} = 1 | \mathbf{X}, \mathbf{A}_{N0})\right].$$

Indexing this probability by N accommodates sequences of degree heterogeneity $\{A_i\}_{i=1}^{N}$ which induce networks with varying degrees of sparsity in the limit. This notation presupposes that the probability distribution of the data may depend on N. It what follows I assume that this dependence reflects a possible relationship between the distribution of the degree heterogeneity and network size (although similar results would hold under more complex sequences of DGPs). For example the distribution of A_i may be such that $\inf_{1 \le i \le N} A_i \to -\infty$

Figure 3: Tetrad isomorphisms



Notes: The $g_{ijkl}(\beta)$ kernel function entering the tetrad logit criterion function (8) is only non-zero when $\{i, j, k, l\}$ belongs to either the two edge ( ), four path ( ), or four cycle ( ) isomorphism class. All 18 subgraphs with non-unique degree sequences are isomorphic to one of these three configurations. If $\{i, j, k, l\}$ belongs to any of the other 8 isomorphism classes depicted above it will not contribute to (8). The degree sequences associated with non-unique subgraphs are $(1, 1, 1, 1)'$, $(2, 2, 2, 2)'$, $(2, 2, 1, 1)'$, $(2, 1, 2, 1)'$, $(2, 1, 1, 2)'$, $(1, 2, 2, 1)'$, $(1, 2, 1, 2)'$ and $(1, 1, 2, 2)'$. The first two are associated with three possible wirings each and the remainder with two wirings each.

as $N \to \infty$, inducing greater sparsity in links as the graph grows in size. The speed at which ρ_N approaches zero is restricted below, and also affects the rate of convergence of $\hat{\beta}_{\text{TL}}$.

Next define

$$T_{ijkl} = \begin{cases} 1, & S_{ij,kl} \in \{-1, 1\} \vee S_{ij,lk} \in \{-1, 1\} \vee S_{ik,lj} \in \{-1, 1\} \\ 0, & \text{otherwise} \end{cases},$$

which equals 1 when the tetrad $\{i, j, k, l\}$ makes a non-zero contribution to the tetrad logit criterion function and zero otherwise. Let

$$\alpha_{q,N} = \Pr\left(T_{i_1 i_2 i_3 i_4} = 1, T_{j_1 j_2 j_3 j_4} = 1\right) \tag{11}$$

be the probability that tetrads $\{i_1, i_2, i_3, i_4\}$ and $\{j_1, j_2, j_3, j_4\}$ both take an identifying configuration when they share $q = 0, 1, 2, 3$ or 4 agents in common. The properties of $\alpha_{q,N}$ as $N \to \infty$ feature in the asymptotic analysis.

The probability that a randomly chosen tetrad contributes to (8) equals $\alpha_{4,N}$. Consistency of $\hat{\beta}_{\text{TL}}$ therefore requires that $\binom{N}{4}\alpha_{4,N} \to \infty$ as $N \to \infty$, ensuring that the number of non-trivial terms entering the criterion function grows large (cf., Chamberlain, 1980, pp. 229 - 230). Under sequences of networks where ρ_N tends toward zero, the probability that a tetrad contributes to (8) is of order ρ_N^2, since any contributing tetrad must contain at least two edges (see Figures 2 and 3). Letting

$$\lambda_N = (N-1)\,\rho_N \tag{12}$$

denote average degree we have that $\binom{N}{4}\alpha_{4,N} = O\left(N^2\lambda_N^2\right)$; consistency of $\hat{\beta}_{\mathrm{TL}}$ for β_0 therefore requires that $N\lambda_N \to \infty$ as N grows large. This condition holds even in the sparse graph case where $\lambda_N = O(1)$. For technical reasons, I also assume that $\lambda_N = \Omega(1)$ (i.e., $\lambda_N \geq \lambda_0 > 0$ for large enough N). Assuming that average degree is bounded from below by a positive constant implies that the network is not empty in the limit.

An explicit identification condition is also needed: the (normalized) Hessian matrix of (8), $\alpha_{4,N}^{-1}\nabla_{\beta\beta}L_N\left(\hat{\beta}_{\mathrm{TL}}\right)$ must converge to a finite non-singular matrix.

Assumption 4. (Conditional Fixed Effects Identification)
(i) $N\lambda_N \to \infty$ *as* $N \to \infty$,
(ii) $\lambda_N = \Omega(1)$,
(iii) $\Gamma_0 = \lim_{N\to\infty} \alpha_{4,N}^{-1}\nabla_{\beta\beta}L_N(\beta_0)$ *is a finite non-singular matrix.*

Part (i) of Assumption 4, which ensures the number of identifying tetrads grows with the order of the network, accommodates degree heterogeneity sequences that induce both sparse, $\lambda_N = O(1)$, and dense $\lambda_N = O(N)$, limiting networks. The rate of convergence of $\hat{\beta}_{\mathrm{TL}}$ to β_0 varies across these two regimes, as detailed further below. Part (ii), as noted above, ensures that the network is non-empty in the limit. To better understand part (iii) of Assumption 4, observe that the scaled Hessian matrix consists of three terms, each of the form

$$\frac{1}{3\alpha_{4,N}}\binom{N}{4}^{-1}\sum_{i<j<k<l}\left|S_{ij,kl}\right| q_{ij,kl}\left(\hat{\beta}_{\mathrm{TL}}\right)\left(1-q_{ij,kl}\left(\hat{\beta}_{\mathrm{TL}}\right)\right)\tilde{W}_{ij,kl}\tilde{W}_{ij,kl}' \tag{13}$$

for the three non-redundant index permutations appearing in (10) above,

$$q_{ij,kl}(\beta) = \left[1+\exp\left(-\tilde{W}_{ij,kl}'\beta\right)\right]^{-1}$$

and $q_{ij,kl} = q_{ij,kl}(\beta_0)$. If we apply iterated expectations to (13) and assume that $\hat{\beta}_{\mathrm{TL}} \xrightarrow{p} \beta_0$ and the distribution of $\tilde{W}_{ij,kl}$ given $S_{ij,kl} = \{-1,1\}$ tends to some limit, we get

$$\Gamma_0 = \mathbb{E}\left[q_{ij,kl}\left(1-q_{ij,kl}\right)\tilde{W}_{ij,kl}\tilde{W}_{ij,kl}'\right| S_{ij,kl} = \{-1,1\}\Big]\Pr\left(S_{ij,kl} = \{-1,1\}\right|T_{ijkl}=1)$$

which suggests the more familiar requirement for identification in binary choice models that $\mathbb{E}\left[\left.\tilde{W}_{ij,kl}\tilde{W}_{ij,kl}'\right| S_{ij,kl}=\{-1,1\}\right]$ is a finite non-singular matrix (e.g., Amemiya, 1985, p. 270). Here expectations are with respect to the limiting conditional distribution of $\tilde{W}_{ij,kl}$ given $S_{ij,kl}=\{-1,1\}$.

Assumptions 1 through 4 are sufficient to prove consistency. They also suffice for showing asymptotic normality, however the argument is more involved. The key challenge in demonstrating normality involves characterizing the sampling properties of the score vector (9). Let U_N equal (9) evaluated at β_0. I begin with the observation that U_N, while not formally a U-Statistic, shares many similarities with one. This suggests a three step approach to demonstrating asymptotic normality. While these steps parallel textbook demonstrations of asymptotic normality of U-Statistics, additional complications arise at each stage due to the more complex structure of dependence across the summand in (9) and because U_N exhibits degeneracy of order 1. First I calculate the variance of U_N using Hoeffding (1948) type arguments. Second I show that the statistic $\sqrt{n\alpha_{2,N}^{-1}}U_N$ is asymptotically equivalent to a Hajek-type projection. The projection takes the form of a summation across dyads *not* agents. Consequently it does not consist of independent components. It does have a special conditional independence structure which I exploit to verify that it, appropriated studentized, obeys a CLT.

To state Theorem 1 I need two more pieces of notation. First, define the projection

$$\bar{s}_{ij}(\beta) \stackrel{\text{def}}{\equiv} \mathbb{E}\left[\left. s_{ijkl}(\beta)\right| X_i, X_j, A_i, A_j, U_{ij}\right].$$

Second, I require an index notation for dyads. Recall that $i=1,2,\ldots$ indexes the N sampled *agents*. Let the boldface indices $\mathbf{i}=\mathbf{1},\mathbf{2},\ldots$ index the $n=\binom{N}{2}$ *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, $D_{\mathbf{i}}=D_{i_1 i_2}$.

Theorem 1. (Large Sample Properties of $\hat{\beta}_{\text{TL}}$) *Under Assumptions 1, 2, 3 and 4:*
(i) $\hat{\beta}_{\text{TL}} \xrightarrow{p} \beta_0$
(ii) $\dfrac{\alpha_{4,N}\sqrt{n\alpha_{2,N}^{-1}}c'\left(\hat{\beta}_{\text{TL}}-\beta_0\right)}{\sqrt{c'\Gamma_0^{-1}\tilde{\Delta}_N\Gamma_0^{-1}c}} \xrightarrow{D} \mathcal{N}(0,36)$
for any $K\times 1$ *vector of real constants* c, $\tilde{\Delta}_N=\frac{1}{n}\sum_{\mathbf{i}=1}^{n}\tilde{\Delta}_{\mathbf{i}}$, *and* $\tilde{\Delta}_{\mathbf{i}}=\alpha_{2,N}^{-1}\mathbb{E}\left[\left.\bar{s}_{\mathbf{i}}\bar{s}_{\mathbf{i}}'\right| X_{i_1},X_{i_2},A_{i_1},A_{i_2}\right]$.

Proof. See Appendix A. □

Theorem 1 follows from the asymptotically linear representation (see equation (35) of Appendix A):

$$\sqrt{n\alpha_{2,N}^{-1}}\alpha_{4,N}\left(\hat{\beta}_{\text{TL}}-\beta_0\right)=-6\Gamma_0^{-1}\left[\frac{1}{\sqrt{n\alpha_{2,N}}}\sum_{i<j}^{N}\bar{s}_{ij}\left(\beta_0\right)\right]+o_p\left(1\right). \tag{14}$$

The components of the sum in (14) are not independent of one another, however they are conditionally independent of each other given $\mathbf{X}$ and $\mathbf{A}$. This conditional independence structure is used to show asymptotic normality by adapting an argument used by Chatterjee (2006) to prove his Theorem 1.1 (which is stated in a form which is not directly applicable here).[11]

A somewhat more evocative, albeit heuristic, way of presenting the result follows by defining

$$\Omega_q=\mathbb{E}\left[s_{i_1i_2i_3i_4}\left(\beta_0\right)s_{j_1j_2j_3j_4}\left(\beta_0\right)'\right|T_{i_1i_2i_3i_4}=1,T_{j_1j_2j_3j_4}=1\right]$$

to equal the conditional covariance of the score summands $s_{i_1i_2i_3i_4}\left(\beta_0\right)$ and $s_{j_1j_2j_3j_4}\left(\beta_0\right)$ when the 4-tuples $\{i_1,i_2,i_3,i_4\}$ and $\{j_1,j_2,j_3,j_4\}$ have $q=0,1,2,3,4$ indices in common and both tetrads take an identifying configuration. If $\lim_{N\to\infty}\tilde{\Delta}_N=\Omega_2$ I get

$$\sqrt{n\alpha_{2,N}^{-1}}\alpha_{4,N}\left(\hat{\beta}_{\text{TL}}-\beta_0\right)\xrightarrow{D}\mathcal{N}\left(0,36\Gamma_0^{-1}\Omega_2\Gamma_0^{-1}\right).$$

In Appendix A I show that the probability that any two tetrads, sharing two agents in common, both take an identifying configuration, $\alpha_{2,N}$, is of order ρ_N^3. The probability that a random tetrad takes an identifying configuration, $\alpha_{4,N}$, is of order ρ_N^2. These two results give $\sqrt{n\alpha_{2,N}^{-1}}\alpha_{4,N}=O(\sqrt{n}\rho_N^{-3/2}\rho_N^{4/2})=O(\sqrt{n\rho_N})=O(\sqrt{N\lambda_N})$ from which a Corollary to Theorem 1 follows.

Corollary 1. (Rate of Convergence) *Under Assumptions 1, 2, 3 and 4, if* $\lim_{N\to\infty}\tilde{\Delta}_N=\Omega_2$ *, then* $\hat{\beta}_{\text{TL}}\xrightarrow{p}\beta_0$ *at rate*
(i) Dense case: $n^{-1/2}$ *if* $\rho_N\to\rho_0>0$ *as* $N\to\infty$*;*
(ii) Sparse Case: $n^{-1/4}$ *if* $\lambda_N=(N-1)\rho_N\to\lambda_0>0$ *as* $N\to\infty$*.*

Under dense graph sequences $\hat{\beta}_{\text{TL}}$ convergences at the usual parametric rate (Recall that the likelihood consists of $n=\binom{N}{2}$ conditionally independent components and hence $\sqrt{n}=\sqrt{\binom{N}{2}}$ is the usual rate). When average density tends toward zero as the graph grows large, the rate of convergence slows. In the sparse case, corresponding to a bounded average degree in

[11] Jochmans (2016b), in an extension of the results presented here to directed networks, uses a CLT due to Rao (2009).

the limit, the rate of convergence is $n^{-1/4}$, which is considerably slower than the parametric rate.[12]

Empirical implementation

From the vantage of an empirical researcher, estimation and inference proceed identically in the sparse and dense cases. Specifically, $\hat{\beta}_{\mathrm{TL}}$ can be calculated using a conventional logit estimation program:

1. For all $\binom{N}{4}$ sampled tetrads calculate $S_{\pi_i\pi_j,\pi_k\pi_l}$ and $\tilde{W}_{\pi_i\pi_j,\pi_k\pi_l}$ for all three non-redundant permutations of the agent-level indices entering kernel (10);

2. Stack these three replicates on top of one another, generating a dataset with $3\binom{N}{4}$ rows and $1+K$ columns;

3. Drop all rows with $S_{\pi_i\pi_j,\pi_k\pi_l} = 0$;

4. Use the retained rows to compute the logit fit of $\mathbf{1}\left(S_{\pi_i\pi_j,\pi_k\pi_l} = 1\right)$ onto $\tilde{W}_{\pi_i\pi_j,\pi_k\pi_l}$. The coefficient on $\tilde{W}_{\pi_i\pi_j,\pi_k\pi_l}$ equals $\hat{\beta}_{\mathrm{TL}}$.

Inference can be based upon the approximation

$$\hat{\beta}_{\mathrm{TL}} \overset{\text{approx}}{\sim} \mathcal{N}\left(\beta_0, \frac{36}{n}\hat{H}^{-1}\hat{\Delta}_{2,N}\hat{H}^{-1}\right),$$

where

$$\hat{H} = \binom{N}{4}^{-1} \sum_{i<j<k<l}^{N} \frac{\partial^2 g_{ijkl}\left(\hat{\beta}_{\mathrm{TL}}\right)}{\partial\beta\partial\beta'}, \quad \hat{\Delta}_{2,N} = \frac{1}{n}\sum_{i<j} \hat{\bar{s}}_{ij}\left(\hat{\beta}_{\mathrm{TL}}\right)\hat{\bar{s}}_{ij}\left(\hat{\beta}_{\mathrm{TL}}\right)'$$

with $\hat{\bar{s}}_{ij}(\beta) = \frac{1}{n-2(N-1)+1}\sum_{k<l,\{i,j\}\cap\{k,l\}=\emptyset} s_{ijkl}(\beta)$. This is the covariance estimator used in the Monte Carlo experiments. Note that $n-2(N-1)+1$ coincides with the $\binom{N-2}{2}$ tetrads that each dyad belongs to.

The actual computation of $\hat{\beta}_{\mathrm{TL}}$ is quite quick, even with medium-sized networks. However the pre-processing of the network data described in steps 1 to 3 above can be computationally expensive. For covariance matrix estimation, the $\hat{H}$ matrix can be recovered from the output of a logit estimation program. The computation of $\hat{\Delta}_{2,N}$ is more expensive. This is because for each of the n dyads an average of $O(n)$ elements must be computed first (for a total of $O(N^4)$ operations in a naive brute force implementation).[13]

[12] Olhede and Wolfe (2014, Theorem 1) find that the mean integrated squared error of their network histogram estimator decays at rate $1/\sqrt{N\lambda_N}$.

[13] The Python code used for the Monte Carlo results reported in the Supplemental Materials incorporates

Joint maximum likelihood (JML) estimation

Let $\mathbf{A}_N$ denote an $N \times 1$ vector of degree heterogeneity values and $\mathbf{A}_{0N}$ the corresponding vector of true values. The N subscript is used in this sub-section where it is helpful to emphasize that the dimension of the incidental parameter vector grows with the sample size. For what follows it is also convenient to define the notation

$$p_{ij}\left(\beta, A_i, A_j\right) \overset{def}{=} \frac{\exp\left(W_{ij}'\beta + A_i + A_j\right)}{1+\exp\left(W_{ij}'\beta + A_i + A_j\right)}.$$

The joint maximum likelihood estimator chooses $\hat{\beta}_{\text{JML}}$ and $\hat{\mathbf{A}}_N$ simultaneously in order to maximize the log-likelihood function

$$l_N\left(\beta, \mathbf{A}_N\right) = \sum_{i<j} \left\{D_{ij} \ln p_{ij}\left(\beta, A_i, A_j\right) + \left(1 - D_{ij}\right) \ln\left[1 - p_{ij}\left(\beta, A_i, A_j\right)\right]\right\}. \tag{15}$$

Some insight in $\hat{\beta}_{\text{JML}}$ is provided by outlining a method of computation. For this purpose it is convenient to note that $\hat{\beta}_{\text{JML}}$ also maximizes the concentrated likelihood

$$l_N^c\left(\beta, \hat{\mathbf{A}}\left(\beta\right)\right) = \sum_{i=1}^{N}\sum_{j<i} D_{ij}\left(W_{ij}'\beta + T_{ij}'\hat{\mathbf{A}}_N\left(\beta\right)\right) - \ln\left[1+\exp\left(W_{ij}'\beta + T_{ij}'\hat{\mathbf{A}}_N\left(\beta\right)\right)\right] \tag{16}$$

where $\hat{\mathbf{A}}_N\left(\beta\right) = \arg\max_{\mathbf{A}\subset\mathbb{R}^N} l_N\left(\beta, \mathbf{A}\right).$

By adapting Theorem 1.5 of Chatterjee et al. (2011) I show that $\hat{\mathbf{A}}_N\left(\beta\right)$, when it exists, is the unique solution to the fixed point problem

$$\hat{\mathbf{A}_N}\left(\beta\right) = \varphi\left(\hat{\mathbf{A}}_N\left(\beta\right)\right) \tag{17}$$

where

$$\varphi\left(\mathbf{A}\right) \overset{def}{=} \begin{pmatrix} \ln D_{1+} - \ln r_1\left(\beta, \mathbf{A}, \mathbf{W}_1\right) \\ \vdots \\ \ln D_{N+} - \ln r_N\left(\beta, \mathbf{A}, \mathbf{W}_N\right) \end{pmatrix}, \tag{18}$$

a number of computational speed-ups by keeping careful track of non-contributing tetrads as estimation proceeds (hence calculations which are known to be zero are omitted). Nevertheless to use tetrad logit on large graphs would require parallelization and/or approximation. Bhattacharya and Bickel (2015) introduce a subsampling procedure for subgraph counts, which might be adapted to produce an estimate of $\hat{\Delta}_{2,N}$ based on subsampling.

with $\mathbf{W}_i = \left(W_{i1}, \ldots, W_{i(i-1)}, W_{i(i+1)}, , W_{iN}\right)'$ and

$$r_i\left(\beta, \mathbf{A}\left(\beta\right), \mathbf{W}_i\right) = \sum_{j \neq i} \frac{\exp\left(W_{ij}'\beta\right)}{\exp\left(-A_j\left(\beta\right)\right) + \exp\left(W_{ij}'\beta + A_i\left(\beta\right)\right)}.$$

That $\hat{\mathbf{A}}_N\left(\beta\right) = \varphi\left(\hat{\mathbf{A}}_N\left(\beta\right)\right)$ can be directly verified by rearranging the sample score of (15). That iteration using (17) converges to $\hat{\mathbf{A}}_N\left(\beta\right) = \arg\max_{\mathbf{A} \in \mathbb{A}^N} l_N\left(\beta, \mathbf{A}\right)$ – when the solution exists – is a direct implication of Lemma 4 in the Supplemental Materials.

The fixed point representation of $\hat{\mathbf{A}}_N\left(\beta\right)$ shows that, while the incidental parameters $\{A_i\}_{i=1}^N$ are agent-specific, their concentrated MLEs are *jointly* determined using all $n = \binom{N}{2}$ dyad observations. To see this observe that if we perturb $\hat{A}_i$, then all values of $\hat{A}_j$ for $i \neq j$ will change. This differs from joint fixed effects estimation in a nonlinear panel data model without time effects. In those models, conditional on the common parameter, the value of $\hat{A}_i\left(\beta\right)$ is a function of only the T observations specific to unit i (e.g., Hahn and Newey, 2004; Arellano and Hahn, 2007). The joint determination of the components of $\hat{\mathbf{A}}_N\left(\beta\right)$ is a direct consequence of the multi-agent nature of the network formation problem and complicates the asymptotic analysis of $\hat{\beta}_{\text{JML}}$.

To characterize the large sample properties of the JML estimates I require some additional notation and an identification condition. It is useful to begin by observing that the population problem is

$$\max_{b \in \mathbb{B}, \mathbf{a}_N \in \mathbb{A}^N} \mathbb{E}\left[\left. l_N\left(b, \mathbf{a}_N\right)\right| \mathbf{X}, \mathbf{A}_{0N}\right],$$

where it is easy to show that

$$\mathbb{E}\left[\left. l_N\left(\beta, \mathbf{A}_N\right)\right| \mathbf{X}, \mathbf{A}_{0N}\right] = -\sum_{i<j} D_{KL}\left(p_{ij} \| \, p_{ij}\left(\beta, A_i, A_j\right)\right) - \sum_{i<j} \mathbf{S}\left(p_{ij}\right),$$

where $D_{KL}\left(p_{ij} \| \, p_{ij}\left(\beta, A_i, A_j\right)\right)$ is the Kullback-Leibler divergence of $p_{ij}\left(\beta, A_i, A_j\right)$ from $p_{ij} \stackrel{def}{=} p_{ij}\left(\beta_0, A_{i0}, A_{j0}\right)$ and $\mathbf{S}\left(p_{ij}\right)$ is the binary entropy function. It is clear that $\left(\beta_0, \mathbf{A}_{0N}\right)$ is *a* maximizer of the population criterion function. The following assumption ensures that it is the *unique* maximizer (and also that this maximizer exists for large enough N).

Assumption 5. (Joint FE Identification)
(i) For $i = 1, 2, \ldots$ the support of A_{i0} is $\mathbb{A}$, a compact subset of $\mathbb{R}$.
(ii) $\mathbb{E}\left[\left. l_N\left(b, \mathbf{a}_N\right)\right| \mathbf{X}, \mathbf{A}_{0N}\right]$ *is uniquely maximized at* $b = \beta_0$ *and* $\mathbf{a}_N = \mathbf{A}_{0N}$ *for large enough* N.

Part (i) of the assumption implies, in combination with Assumption 2, that

$$p_{ij}\left(\beta, A_i, A_j\right) \in (\kappa, 1-\kappa) \tag{19}$$

for some $0 < \kappa < 1$ and for all $(A_i, A_j) \in \mathbb{A} \times \mathbb{A}$ and $\beta \in \mathbb{B}$. Condition (19) implies that in large networks the number of observed links per agent will be proportional to the number of sampled agents. Put differently it implies a dense sequence of graphs. It might be possible to relax part (i) to accommodate sequences of $\{A_{0i}\}_{i=1}^{N}$ that diverge at some (slow enough) rate (e.g., $\sup_{1\leq i\leq N} |A_{0i}| = O(\log\log N)$), but the structure of the proofs of Theorems 2, 3 and 4 suggest that any feasible sequence will still result in a non-sparse graph (i.e., agents will have a large number of links in the limit).[14] This contrasts sharply with the tetrad logit estimator, where consistency under sparse graph sequences was established.

Part (ii) of Assumption 5 is an identification condition. It will generally hold if there is sufficient variance in each column of $\mathbf{W}_i = \left(W_{i1}, \ldots, W_{i(i-1)}, W_{i(i+1)}, , W_{iN}\right)'$.

The first theorem establishes consistency of $\hat{\beta}_{\mathrm{JML}}$.

Theorem 2. *Under Assumptions 1, 2, 3 and 5*

$$\hat{\beta}_{\mathrm{JML}} \xrightarrow{p} \beta_0.$$

Proof. See the Supplemental Materials. □

A simple intuition for Theorem 2 is as follows. Rearranging the likelihood yields

$$\begin{aligned} l_N\left(\beta, \mathbf{A}_N\right) &= \sum_{i<j} \left(D_{ij} - p_{ij}\right) \ln\left(\frac{p_{ij}\left(\beta, A_i, A_j\right)}{1 - p_{ij}\left(\beta, A_i, A_j\right)}\right) \\ &\quad - \sum_{i<j} D_{KL}\left(p_{ij} \| \, p_{ij}\left(\beta, A_i, A_j\right)\right) - \sum_{i<j} \mathbf{S}\left(p_{ij}\right) \end{aligned} \tag{20}$$

An implication of (19) is that $\left(D_{ij} - p_{ij}\right) \ln\left(\frac{p_{ij}(\beta, A_i, A_j)}{1-p_{ij}(\beta, A_i, A_j)}\right)$ is a bounded random variable. This fact and Hoeffding's (1963) inequality can be used to show that the first component of $n^{-1} l_N\left(\beta, \mathbf{A}\right)$ is $o_p(1)$ uniformly in $\beta \in \mathbb{B}$ and $\mathbf{A}_N \in \mathbb{A}^N$. The last term in (20) is constant in β. In large samples a maximizer of $l_N\left(\beta, \mathbf{A}\right)$ will therefore be close to a minimizer of the sum of the n Kullback-Leibler measures of divergence of $p_{ij}\left(\beta, A_i, A_j\right)$ from p_{ij} across all dyads. From part (ii) of Assumption 5 this minimizer is unique.[15]

---

[14] In a recent working paper, Yan et al. (2016) explore such an extension for a directed network analog of the model studied here.

[15] The argument is close to that of a standard M-estimator consistency proof (e.g., Amemiya 1985, pp.

A more involved argument shows that it is possible to estimates the elements of $\mathbf{A}_{0N}$ with uniform accuracy.

Theorem 3. *With probability* $1 - O\left(N^{-2}\right)$

$$\sup_{1 \leq i \leq N} \left| \hat{A}_i - A_{i0} \right| < O\left(\sqrt{\frac{\ln N}{N}} \right).$$

Proof. See the Supplemental Materials. □

Chatterjee et al. (2011) show uniform consistency of $\hat{A}_i$ in the model with no dyad-level covariates. Theorem 3 follows from a combination of Theorem 2 above and an adaptation of their results. It is also closely related to Simons and Yao's (1999) analysis of the Bradley-Terry model of paired comparisons.

The key intuition is as follows. Under dense graph sequences we effectively observe $N - 1$ linking decisions per agent. That is we observe whether agent i links with j for all $j \neq i$. This feature of the problem allows for consistent estimation of A_{i0} for each agent. The argument is complicated by the fact that agents $i's$ and agent $j's$ sequences of link decisions are dependent. However this dependence is weak, only arising via the presence of D_{ij} in both link sequences.[16]

Establishing asymptotic normality of $\hat{\beta}_{\mathrm{JML}}$ is also involved. This is because the sampling properties of $\hat{\beta}_{\mathrm{JML}}$ are influenced by the estimation error in $\hat{\mathbf{A}}_N$. This influence generates bias in the limit distribution of $\hat{\beta}_{\mathrm{JML}}$. This bias is similar to that which arises in large-N, large-T joint fixed effects estimation of nonlinear panel data models (Hahn and Newey, 2004; Arellano and Hahn, 2007). An additional challenge here, not present in the panel data problem, is to characterize the probability limit of the (suitably normalized) Hessian matrix of the concentrated log-likelihood. This matrix depends on the inverse of the $N \times N$ block of the full likelihood's Hessian associated with the incidental parameters. This sub-matrix, unlike in the corresponding panel data problem, is not diagonal due to the weak dependence across different agents' link sequences. The inverse of this sub-matrix is not available in closed form and hence must be approximated.[17]

---

106 - 107). The presence of the incidental parameters $\{A_i\}_{i=1}^N$ complicates the argument. This handled by "concentrating them out" of the problem.

[16]Lemma 6 in the Supplemental Materials additionally establishes asymptotic normality of any sub-vector of $\hat{\mathbf{A}}$ of fixed length:

$$\sqrt{N}\left(\hat{\mathbf{A}} - \mathbf{A}\right)_{1:L} \xrightarrow{D} \mathcal{N}\left(0, \operatorname{diag}\left(\frac{1}{\mathbb{E}\left[p_{1j}\left(1 - p_{1j}\right)\right]}, \ldots, \frac{1}{\mathbb{E}\left[p_{Lj}\left(1 - p_{Lj}\right)\right]}\right)\right).$$

[17]In the proof I use some matrix approximation results originally developed in the context of the Bradley-

To state the form of the limit distribution define

$$\mathcal{I}_0(\beta) = \lim_{N\to\infty} -\binom{N}{2}^{-1} \frac{\partial^2 l_N^c\left(\beta_0, \hat{\mathbf{A}}(\beta_0)\right)}{\partial\beta\partial\beta'}, \tag{21}$$

and also

$$B_0 = -\lim_{N\to\infty} \frac{1}{2\sqrt{n}} \sum_{i=1}^{N} \frac{\frac{1}{N-1}\sum_{j\neq i} p_{ij}(1-p_{ij})(1-2p_{ij}) W_{ij}}{\frac{1}{N-1}\sum_{j\neq i} p_{ij}(1-p_{ij})}. \tag{22}$$

Theorem 4. *Under Assumptions 1, 2, 3 and 5*[18]
(i) $\hat{\beta}_{\text{JML}} = \beta_0 + \frac{\mathcal{I}_0^{-1}(\beta)B_0}{\sqrt{n}} + o_p(1)$
(ii) $\frac{\sqrt{n}c'\left(\hat{\beta}_{\text{JML}}-\beta_0\right)-c'\mathcal{I}_0^{-1}(\beta)B_0}{\left(c'\mathcal{I}_0^{-1}(\beta)\mathcal{I}_N(\beta)\mathcal{I}_0^{-1}(\beta)c\right)^{1/2}} \xrightarrow{D} \mathcal{N}(0,1)$,
or any $K\times 1$ *vector of real constants* c *and* $\mathcal{I}_N(\beta)$ *as defined in the Supplemental Materials.*

Proof. See the Supplemental Materials. □

Empirical implementation

Computation of $\hat{\beta}_{\text{JML}}$ is possible by computing the logit fit of D_{ij} onto W_{ij} and T_{ij}. The dimension of the latter vector is N, and hence the concentration approach outlined above will be more reliable in practice. For small and/or sparse networks $\hat{\beta}_{\text{JML}}$ may not exist. Inference, noting that $\mathcal{I}_N(\beta) \xrightarrow{p} \mathcal{I}_0(\beta)$, may be based on the approximation

$$\hat{\beta}_{\text{JML}} \overset{approx}{\sim} \mathcal{N}\left(\beta_0 + \frac{\mathcal{I}_0^{-1}(\beta)B_0}{\sqrt{n}}, \frac{\mathcal{I}_0^{-1}(\beta)}{n}\right).$$

Hence the standard errors reported by a conventional logit command will be valid (alternatively a sandwich estimator may be used).

Although conventional logit standard errors will be valid, confidence intervals computed using them with not be, due to the bias in the limit distribution. Consequently, for inference it is important to bias-correct $\hat{\beta}_{\text{JML}}$. There are many possible approaches to bias correction (cf., Hahn and Newey, 2004; Fernández-Val and Weidner, 2016). I use the iterated bias correction procedure outlined in Hahn and Newey (2004) in the Monte Carlo experiments summarized in the Supplemental Materials. In this procedure $\hat{\beta}_{\text{JML}}$ is used to replace β_0 in the sample

Terry model for paired comparisons (cf., Simons and Yao, 1998). Fernández-Val and Weidner (2016) encounter a related problem in their extension of Hahn and Newey (2004) to include time effects.

[18]To relate Theorem 4 to analogous results from the large-N, large-T non-linear panel data literature observe that for each agent we observe $N-1$ linking decisions; corresponding to "T" in the panel data case. The bias term is thus $O(1/N) = O(1/\sqrt{n})$, analogous to the $O(1/T)$ bias term in the panel data case (e.g., Hahn and Newey (2004)).

analogs of (21) and (22), yielding $\hat{\mathcal{I}}_1$ and $\hat{B}_1$. Next compute $\hat{\beta}_{\text{BC},1}$ as $\hat{\beta}_{\text{BC},1} = \hat{\beta}_{\text{JML}} - \frac{\hat{\mathcal{I}}_1^{-1}\hat{B}_1}{\sqrt{n}}$. Plug this estimate of β_0 back into (21) and (22) and compute $\hat{\beta}_{\text{BC},2} = \hat{\beta}_{\text{BC},1} - \frac{\hat{\mathcal{I}}_2^{-1}\hat{B}_2}{\sqrt{n}}$. Repeat until $\hat{\beta}_{\text{BC},b} = \hat{\beta}_{\text{BC},b+1} \stackrel{def}{=} \hat{\beta}_{\text{BC}}$. In principle the limiting variance of $\sqrt{n}\left(\hat{\beta}_{\text{BC}} - \beta_0\right)$ need not coincide with the one given in Theorem 4, although the results of Hahn and Newey (2004) and others suggest it should.

3 Areas for further work

As noted in the introduction, one limitation of the model studied here is that it excludes interdependencies in link preferences. This omission raises two natural questions. First, can one construct a test for the assumption of no interdependencies in link formation? Second, can one augment the model to include such interdependencies?

Consider the testing problem first. A natural way to include interdependencies in preferences is to posit that links form according to

$$D_{ij} = \mathbf{1}\left(\delta_0\left(\sum_{k=1}^{N} D_{ik}D_{jk}\right) + W_{ij}'\beta_0 + A_i + A_j - U_{ijt} \geq 0\right) \tag{23}$$

so that an ij link is more likely if i and j share many friends in common. Transitivity in link structure is predicted by many models of strategic network formation (see Graham (2015) and de Paula (2016) for discussion and references). Link rule (23) results in an incomplete model of network formation: for a given draw of $\mathbf{U}$ there will generally be multiple equilibrium network configurations consistent with (23) (cf., Tamer, 2003). However, under the null of $\delta_0 = 0$ the model coincides with the one analyzed here, which suggests a Score/LM test for neglected transitivity (cf., Hahn et al., 2016). The TL estimator may be especially convenient for this purposes, since its "score" vector does not depend on $\{A_i\}_{i=1}^N$. The study of this approach to specification testing (and other approaches) would be an interesting topic for future research. Dzemski (2014) develops a different approach to testing for neglected transitivity.

Turning to the second question, if the econometrician observes a network for two periods, then the incorporation of interdependencies in link formation, albeit of a particular kind, is possible. Assume that individuals i and j form a period t link, for $i = 1, \ldots, N$ and $j < i$, according to the rule, for example,

$$D_{ijt} = \mathbf{1}\left(\gamma D_{ijt-1} + \delta\sum_{k=1}^{N} D_{ikt-1}D_{jkt-1} + \left(W_{ijt}^*\right)'\beta^* + A_i + A_j - U_{ijt} \geq 0\right), \tag{24}$$

where U_{ijt} is iid across pairs and over time as well as logistic. This model allows the probability of a period t ij link to depend on (i) whether i and j were linked in the prior period and (ii) on the number of friends they shared in common in the prior period. It incorporates both state-dependence and a taste for transitivity in links.

In the two period case ($t = 0, 1$), both the tetrad logit and joint maximum likelihood estimates remain valid, with outcome $D_{ij} = D_{ij1}$, regressor vector $W_{ij} = \left(D_{ij0}, \sum_{k=1}^{N} D_{ik0}D_{jk0}, \left(W_{ij1}^{*}\right)'\right)'$, and coefficient vector $\beta = \left(\gamma, \delta, (\beta^{*})'\right)'$ (see Nadler (2015) for a closely related empirical illustration). This observation hinges critically on the way in which agent-level heterogeneity is modeled. For example, the conditional estimator is based on within-*agent* variation in a given network; over time contrasts are not used. If $A_i + A_j$ were replaced with, say, $A_{ij} = B_i + B_j + h(C_i, C_j)$ for B_i and C_i agent-specific heterogeneity and $h(\bullet, \bullet)$ symmetric but otherwise arbitrary, then identification of β would need to rely on (over-time) within-*dyad* variation and a variant of the initial condition problem that occurs in single agent dynamic panel data analysis would arise. Graham (2013; 2016) studies models of this type.

A Tetrad logit

A.1 Part 1: Consistency

To show that $\hat{\beta}_{\text{TL}} \xrightarrow{p} \beta_0$ I verify the conditions of Theorem 4.1.1 in Amemiya (1985, pp. 106 - 107) (see also Theorem 2.1 of Newey and McFadden (1994)). The main difficulty is finding a normalization of the objective function that can accommodate sequences of data generating processes where $\rho_N \to 0$ as $N \to \infty$ (cf., Newey and McFadden, 1994, p. 2123). Recall that $g_{ijkl}(\beta) = l_{ij,kl}(\beta) + l_{ij,lk}(\beta) + l_{ik,lj}(\beta)$ for $l_{ij,kl}(\beta)$ as defined in (6). The expected value of the tetrad logit criterion function, normalized by $\alpha_{4,N}$, therefore equals

$$\mathbb{E}\left[\frac{L_N(\beta)}{\alpha_{4,N}}\right] = \frac{1}{3\Pr(T_{ijkl} = 1)}$$
$$\{\Pr(S_{ij,kl} \in \{-1,1\})\,\mathbb{E}\left[l_{ij,kl}(\beta) \middle| S_{ij,kl} \in \{-1,1\}\right]$$
$$+\Pr(S_{ij,lk} \in \{-1,1\})\,\mathbb{E}\left[l_{ij,lk}(\beta) \middle| S_{ij,lk} \in \{-1,1\}\right]$$
$$+\Pr(S_{ik,lj} \in \{-1,1\})\,\mathbb{E}\left[l_{ik,lj}(\beta) \middle| S_{ik,lj} \in \{-1,1\}\right]\}$$

By exchangeability each of the three terms to the right of the equality in the expression above are equal to one another. We may therefore consider only the first without loss of generality. Recall that $q_{ij,kl}(\beta) = \left[1 + \exp\left(-\tilde{W}'_{ij,kl}\beta\right)\right]^{-1}$, $q_{ij,kl} = q(\beta_0)$ and define

$$Q(\beta) = -\mathbb{E}\left[D_{\mathrm{KL}}\left(q_{ij,kl} \| q_{ij,kl}(\beta)\right) + \mathbf{S}(q_{ij,kl}) \middle| S_{ij,kl} \in \{-1,1\}\right],$$

where $D_{\mathrm{KL}}\left(q_{ij,kl} \| q_{ij,kl}(\beta)\right)$ is the Kullback-Leibler divergence of $q_{ij,kl}(\beta)$ from $q_{ij,kl}$, $\mathbf{S}(q_{ij,kl})$ is the binary entropy function, and the expectation is with respect to the limiting conditional distribution of $\tilde{W}_{ij,kl}$ given $S_{ij,kl} \in \{-1,1\}$. Using (6) and the equality

$$\Pr\left(S_{ij,kl} = \{-1,1\} \middle| T_{ijkl} = 1\right) = \frac{\Pr\left(S_{ij,kl} = \{-1,1\}\right)}{\Pr\left(T_{ijkl} = 1\right)},$$

I get

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{L_N(\beta)}{\alpha_{4,N}}\right] = Q(\beta)\Pr\left(S_{ij,kl} = \{-1,1\} \middle| T_{ijkl} = 1\right) \qquad (25)$$
$$Q(\beta) \cdot O(1)$$

since $\Pr\left(S_{ij,kl} \in \{-1,1\}\right)$ and $\Pr\left(T_{ijkl} = 1\right)$ are of the same order.

By the properties of the Kullback-Leibler divergence, β_0 is a maximizer of $Q(\beta)$. Uniqueness of this maximum follows from part (iii) of Assumption 4, which implies global concavity of $Q(\beta)$ in β. Consistency of $\hat{\beta}_{\mathrm{TL}}$ for β_0 then follows if

$$\alpha_{4,N}^{-1} L_N(\beta) \xrightarrow{p} Q(\beta)\Pr\left(S_{ij,kl} = \{-1,1\} \middle| T_{ijkl} = 1\right)$$

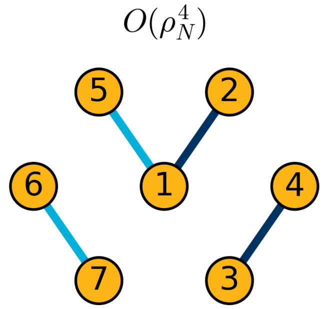
uniformly in $\beta \in \mathbb{B}$.

To show uniform convergence I use a Hoeffding (1948) variance decomposition to calculate

$$\mathbb{V}(L_N(\beta)) = \binom{N}{4}^{-2} \sum_{q=0}^{4} \binom{N}{4}\binom{4}{q}\binom{N-4}{4-q}\xi_{q,N}(\beta) \qquad (26)$$

for $\xi_{q,N}(\beta) = \mathbb{C}\left(g_{i_1 i_2 i_3 i_4}(\beta), g_{j_1 j_2 j_3 j_4}(\beta)\right)$ when the 4-tuples $\{i_1, i_2, i_3, i_4\}$ and $\{j_1, j_2, j_3, j_4\}$ have $q = 0,1,2,3,4$ indices in common. The first, $q = 0$, term in (26) is zero by Assumption 1, which implies that links form independently conditional on $\mathbf{X}$ and $\mathbf{A}$. To determine the order of the remaining terms in (26) I calculate the frequency with which both $g_{i_1 i_2 i_3 i_4}(\beta)$ and $g_{j_1 j_2 j_3 j_4}(\beta)$ are non-zero conditional on the event that they share $q = 1,2,3,4$ indices in common. Recall that, for $\{i_1, i_2, i_3, i_4\}$ and $\{j_1, j_2, j_3, j_4\}$ the index sets associated with two tetrads, $\alpha_{q,N} = \Pr\left(T_{i_1 i_2 i_3 i_4} = 1, T_{j_1 j_2 j_3 j_4} = 1\right)$ equals the probability that both tetrads contribute to the tetrad logit criterion function (8) (when they share $q = 0,1,2,3,4$ agents

Figure 4: Tetrad stitching with $q=1$ agents in common



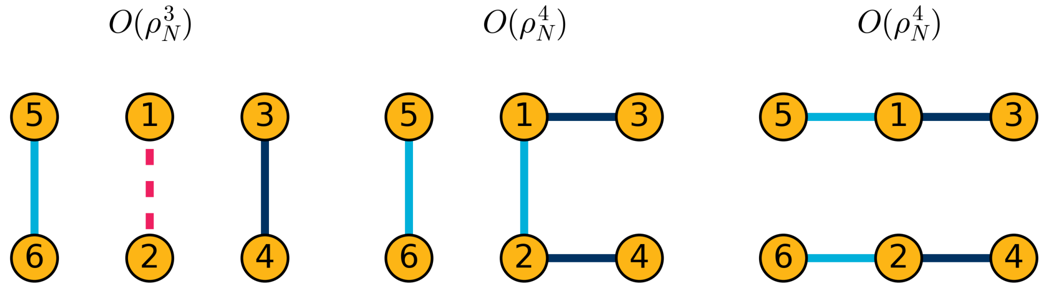
Notes: Depiction of the single non-isomorphic way to join two tetrads, each with degree sequences of $(1,1,1,1)$, such that they share exactly one agent in common. Edges in tetrad $\{1,2,3,4\}$ are depicted with solid dark (Berkeley Blue) lines, while those in tetrad $\{1,5,6,7\}$ are depicted by solid light (Lawrence colored) lines. Any such joining results in a subgraph with four edges. Observe that a joining of a two edge ( ) and, say, a four path ( ), such that they share one node in common, results in a subgraph with five edges. Similar reasoning indicates that $\alpha_{1,N}$ is of order ρ_N^4.

in common). To get the order of $\{\alpha_{q,N}\}_{q=1,2,3,4}$ we need only consider the sparsest identifying tetrad configuration: a two edge ( ), with degree sequence $(1,1,1,1)$. Figure 4 depicts the single way, up to isomorphisms, that two isomorphisms can be joined such that they share a single agent in common. This join requires four edges so that $\xi_{1,N}(\beta) = O(\alpha_{1,N}) = O(\rho_N^4)$. By reference to Figures 5 and 6, we further have that $\xi_{2,N}(\beta) = \xi_{3,N}(\beta) = O(\rho_N^3)$. Finally both $\xi_{4,N}(\beta)$ and $\alpha_{4,N}$ are $O(\rho_N^2)$. These facts give

$$\begin{aligned}\mathbb{V}\left(\frac{L_N(\beta)}{\alpha_{4,N}}\right) &= O\left(N^{-1}\right) + O\left(\rho_N^{-1}N^{-2}\right) + O\left(\rho_N^{-1}N^{-3}\right) + O\left(\rho_N^{-2}N^{-4}\right) \\ &= O\left(\frac{1}{N}\right) + O\left(\frac{1}{N\lambda_N}\right) + O\left(\frac{1}{N^2\lambda_N}\right) + O\left(\frac{1}{N^2\lambda_N^2}\right),\end{aligned}$$

with $\lambda_N = (N-1)\rho_N$. This variance converges to zero as long as $N\lambda_N \to \infty$ as $N \to \infty$ as is asserted by part (i) of Assumption 4. We therefore have that $\alpha_{4,N}^{-1}L_N(\beta)$ convergences in mean square to $Q(\beta)$ at rate $\max\left(\frac{1}{N}, \frac{1}{N\lambda_N}\right)$. By concavity of $L_N(\beta)$ in β this convergence is uniform in $\beta \in \mathbb{B}$. Since conditions A, B and C of Theorem 4.1.1 in Amemiya (1985, pp. 106 - 107) hold, part (i) of the Theorem 1 follows.

Figure 5: Tetrad stitching with $q = 2$ agents in common



Notes: Depiction of the three non-isomorphic way to join two  tetrads, each with degree sequences of $(1, 1, 1, 1)$, such that they share exactly two agents in common. Edges shared by both tetrads are depicted with dashed (Rose Garden colored) lines. Edges in tetrad $\{1, 2, 3, 4\}$ are depicted with solid dark (Berkeley Blue) lines, while those in tetrad $\{1, 2, 5, 6\}$ are depicted by solid light (Lawrence colored) lines. Moving from left-to-right, the first figure shows a joining which requires only three edges in total, one shared by both tetrads. The second and third joinings involve no shared edges and hence each require four edges in total.

A.2 Part II: Asymptotic normality

A Taylor expansion of the first order condition of the tetrad logit criterion function yields, after re-arrangement and re-scaling,

$$\sqrt{n\alpha_{2,N}^{-1}}\alpha_{4,N}\left(\hat{\beta}_{\mathrm{TL}} - \beta_0\right) = -\left[\alpha_{4,N}^{-1}\nabla_{\beta\beta}L_N\left(\bar{\beta}\right)\right]^{+} \times \sqrt{n\alpha_{2,N}^{-1}}U_N$$

with $\bar{\beta}$ a mean value between $\hat{\beta}_{\mathrm{TL}}$ and β_0 which may vary from row to row, the $+$ superscript denoting a Moore-Penrose inverse, and

$$U_N \stackrel{\mathrm{def}}{\equiv} \binom{N}{4}^{-1} \sum_{i<j<k<l} s_{ijkl}\left(\beta_0\right) = 0,$$

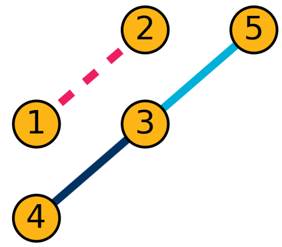
equal to the 'score' vector of the TL criterion function.

For what follows it is helpful to note, once again, that $\alpha_{0,N} = O\left(\rho_N^4\right)$, $\alpha_{1,N} = O\left(\rho_N^4\right)$, $\alpha_{2,N} = O\left(\rho_N^3\right)$, $\alpha_{3,N} = O\left(\rho_N^3\right)$ and $\alpha_{4,N} = O\left(\rho_N^2\right)$ (see Figures 4, 5 and 6). To demonstrate asymptotic normality I proceed by (a) showing that $\left[\alpha_{4,N}^{-1}\nabla_{\beta\beta}L_N\left(\bar{\beta}\right)\right]^{+} \xrightarrow{p} \Gamma_0^{-1}$ for Γ_0 defined in part (iii) of Assumption 4 of the main text, (b) verifying that U_N (scaled by $\sqrt{n\alpha_{2,N}^{-1}}$) obeys a central limit theorem and (c) combining the (a) and (b) to show the final result.

Demonstrating part (a) requires, as in the proof of consistency, verifying that $\alpha_{4,N}^{-1}\binom{N}{4}^{-1}$ is the appropriate scaling of the Hessian matrix (allowing for sparse network sequences). An

Figure 6: Tetrad Stitching ($q = 3$)

$O(\rho_N^3)$



Notes: Depiction of the single non-isomorphic way to join two tetrads, each with degree sequences of $(1, 1, 1, 1)$, such that they share exactly three agents in common. Edges shared by both tetrads are depicted with dashed (Rose Garden colored) lines. Edges in tetrad $\{1, 2, 3, 4\}$ are depicted with solid dark (Berkeley Blue) lines, while those in tetrad $\{1, 2, 3, 5\}$ are depicted by solid light (Lawrence colored) lines.

application of Lemma 2.9 of Newey and McFadden (1994, p. 2138) then gives

$$\sup_{\beta \in \mathbb{B}} \left\| \alpha_{4,N}^{-1} \nabla_{\beta\beta} L_N(\beta) - \Gamma_0 \right\| \xrightarrow{p} 0.$$

Finally, non-singularity of Γ_0 and the continuous mapping theorem, yields the required result. The approach to showing part (b) is as described in the main text immediately before the statement of the Theorem.

Step 1, Hessian convergence

Calculation gives a Hessian matrix of

$$\begin{aligned} \nabla_{\beta\beta} L_N(\beta) = \binom{N}{4}^{-1} \sum_{i<j<k<l} \frac{1}{3} \Big\{ & |S_{ij,kl}| \, q_{ij,kl}(\beta) \left(1 - q_{ij,kl}(\beta)\right) \tilde{W}_{ij,kl} \tilde{W}'_{ij,kl} \\ & + |S_{ij,lk}| \, q_{ij,lk}(\beta) \left(1 - q_{ij,lk}(\beta)\right) \tilde{W}_{ij,lk} \tilde{W}'_{ij,lk} \\ & |S_{ik,lj}| \, q_{ik,lj}(\beta) \left(1 - q_{ik,lj}(\beta)\right) \tilde{W}_{ik,lj} \tilde{W}'_{ik,lj} \Big\}. \end{aligned} \quad (27)$$

Next recall the definition of the $L_{2,1}$ norm:

$$\|\mathbf{A}\|_{2,1} = \sum_{j=1}^{n} \left[\sum_{i=1}^{m} |a_{ij}|^2 \right]^{1/2}. \quad (28)$$

Applying this norm to $\nabla_{\beta\beta}L_N(\beta)-\nabla_{\beta\beta}L_N(\beta_*)$, the mean value inequality, and compatibility of the Fronbenius matrix norm with the Euclidean vector norm, gives

$$\begin{aligned}\left\|\nabla_{\beta\beta}L_N(\beta)-\nabla_{\beta\beta}L_N(\beta_*)\right\|_{2,1} &\leq \sum_{p=1}^{K}\left\|\frac{\partial}{\partial\beta'}\left\{\frac{\partial^2 L_N(\bar{\beta})}{\partial\beta_p\partial\beta}\right\}\right\|_F \|\beta-\beta_*\|_2 \\ &\leq \sum_{p=1}^{K}\binom{N}{4}^{-1}\left\|\sum_{i<j<k<l}\frac{1}{3}\left\{|S_{ij,kl}|\,\tilde{W}_{ij,kl}\tilde{W}'_{ij,kl}\tilde{W}_{p,ij,kl}\right.\right. \\ &\quad \left.\left.+\,|S_{ij,lk}|\,\tilde{W}_{ij,lk}\tilde{W}'_{ij,lk}\tilde{W}_{p,ij,lk}+|S_{ik,lj}|\,\tilde{W}_{ik,lj}\tilde{W}'_{ik,lj}\tilde{W}_{p,ik,lj}\right\}\right\|_F \\ &\quad \times\|\beta-\beta_*\|_2,\end{aligned}$$

with the second inequality following from the fact that

$$\frac{\partial}{\partial\beta'}\left\{\frac{\partial^2 l_{ij,kl}(\beta)}{\partial\beta_p\partial\beta}\right\}=|S_{ij,kl}|\,q_{ij,kl}(\beta)\left(1-q_{ij,kl}(\beta)\right)\left(1-2q_{ij,kl}(\beta)\right)\tilde{W}_{ij,kl}\tilde{W}'_{ij,kl}\tilde{W}_{p,ij,kl},$$

for $p=1,\ldots,K$ and $0\leq|q_{ij,kl}(\beta)(1-q_{ij,kl}(\beta))(1-2q_{ij,kl}(\beta))|\leq 1$. Since the $\mathbb{B}$ and $\mathbb{W}$ are both compact subsets of $\mathbb{R}^K$ (Assumption 2), and $\binom{N}{4}^{-1}\sum_{i<j<k<l}T_{ijkl}=O(\rho_N^2)=O(\alpha_{4,N})$, we have, after re-scaling the Hessian,

$$\left\|\alpha_{4,N}^{-1}\nabla_{\beta\beta}L_N(\beta)-\alpha_{4,N}^{-1}\nabla_{\beta\beta}L_N(\beta_*)\right\|_{2,1}\leq O(1)\|\beta-\beta_*\|_2\,. \tag{29}$$

Condition (29) allows for an application Lemma 2.9 of Newey and McFadden (1994, p. 2138).

Step 2a, Calculating the variance of U_N

For $s_{ijkl}\stackrel{\text{def}}{\equiv}s_{ijkl}(\beta_0)$ let

$$\Delta_{q,N}=\mathbb{C}\left(s_{i_1i_2i_3i_4},s_{j_1j_2j_3j_4}\right) \tag{30}$$

equal the covariance of $s_{i_1i_2i_3i_4}(\beta_0)$ and $s_{j_1j_2j_3j_4}(\beta_0)$ when the 4-tuples $\{i_1,i_2,i_3,i_4\}$ and $\{j_1,j_2,j_3,j_4\}$ have $q=0,1,2,3,4$ indices in common. A Hoeffding (1948) variance decomposition gives

$$\mathbb{V}(U_N)\;=\;\binom{N}{4}^{-1}\binom{4}{2}\binom{N-4}{4-2}\Delta_{2,N}+\binom{N}{4}^{-1}\binom{4}{3}\binom{N-4}{4-3}\Delta_{3,N}+\binom{N}{4}^{-1}\binom{4}{4}\binom{N-4}{4-4}\Delta_{4,N}, \tag{31}$$

where I use the fact that $\mathbb{C}(s_{i_1i_2i_3i_4},s_{j_1j_2j_3j_4})=0$ whenever the tetrads $\{i_1,i_2,i_3,i_4\}$ and $\{j_1,j_2,j_3,j_4\}$ share zero or one agents in common (i.e., $\Delta_{0,N}=\Delta_{1,N}=0$). That $\Delta_{1,N}=0$ indicates that U_N exhibits degeneracy of order 1. To understand this degeneracy consider the ANOVA decomposition

$$\begin{aligned}\mathbb{C}\left(s_{i_1i_2i_3i_4}, s_{j_1j_2j_3j_4}\right) &= \mathbb{E}\left[\mathbb{C}\left(\left. s_{i_1i_2i_3i_4}, s_{j_1j_2j_3j_4}\right| \mathbf{X}, \mathbf{A}\right)\right] + \mathbb{C}\left(\mathbb{E}\left[\left. s_{i_1i_2i_3i_4}\right| \mathbf{X}, \mathbf{A}\right], \mathbb{E}\left[\left. s_{j_1j_2j_3,i_4}\right| \mathbf{X}, \mathbf{A}\right]\right) \\ &= \mathbb{E}\left[\mathbb{C}\left(\left. s_{i_1i_2i_3i_4}, s_{j_1j_2j_3j_4}\right| \mathbf{X}, \mathbf{A}\right)\right]\end{aligned}$$

with the second equality following from the conditional mean zero property of the score function. Next observe that since edges form independently conditional on $\mathbf{X}$ and $\mathbf{A}$, the covariance to the right of the second equality will be zero unless the two tetrads share at least one dyad in common (i.e., $q \geq 2$).

Now consider the case where the two tetrads share two indices (i.e., a dyad) in common (i.e., $\Delta_{2,N}$). Recalling the discussion given in the consistency proof above, we have that $\Delta_{2,N}$ is $O\left(\alpha_{N,2}\right) = O\left(\rho_N^3\right)$. Proceeding analogously we have that $\Delta_{3,N} = O\left(\rho_N^3\right)$ and $\Delta_{4,N} = O\left(\rho_N^2\right)$. Putting these calculations together we have that

$$\begin{aligned}\mathbb{V}\left(\sqrt{n\alpha_{2,N}^{-1}}U_N\right) &\simeq \alpha_{2,N}^{-1}36\Delta_{2,N} + \alpha_{2,N}^{-1}\tbinom{N}{2}\tbinom{N}{3}^{-1}16\Delta_{3,N} + \alpha_{2,N}^{-1}\tbinom{N}{2}\tbinom{N}{4}^{-1}\Delta_{4,N} \\ &= O\left(1\right) + O\left(\frac{1}{N}\right) + O\left(\frac{1}{N\lambda_N}\right)\end{aligned} \tag{32}$$

The first term in (32) dominates if $N\lambda_N \to \infty$ as $N \to \infty$.

Step 2b, Projection:

Consider the link structure of two tetrads, $\{i,j,k,l\}$ and $\{i,j,m,p\}$, sharing the dyad $\{i,j\}$ in common. By virtue of conditional independence of link decisions, and Assumption 3, the configurations of $\{i,j,k,l\}$ and $\{i,j,m,p\}$ are independent conditional on X_i, X_j, A_i, A_j and U_{ij} giving

$$\mathbb{E}\left[\left. s_{ijkl}\left(\beta_0\right) s_{ijmp}\left(\beta_0\right)'\right| X_i, X_j, A_i, A_j, U_{ij}\right] = \bar{s}_{ij}\left(\beta_0\right)\bar{s}_{ij}\left(\beta_0\right)'$$

for $\bar{s}_{ij}\left(\beta_0\right) \stackrel{\text{def}}{\equiv} \mathbb{E}\left[\left. s_{ijkl}\left(\beta_0\right)\right| X_i, X_j, A_i, A_j, U_{ij}\right]$. Iterated expectations then implies that, for $\Delta_{q,N}$ as defined in (30) above, $\Delta_{2,N} = \mathbb{E}\left[\bar{s}_{ij}\left(\beta_0\right)\bar{s}_{ij}\left(\beta_0\right)'\right]$. Projecting U_N onto an arbitrary function of $(X_i, X_j, A_i, A_j, U_{ij})$ then yields

$$\mathbb{E}\left[\left. U_N\right| X_i, X_j, A_i, A_j, U_{ij}\right] = \binom{N}{4}^{-1}\sum_{k_1<k_2<k_3<k_3}\mathbb{E}\left[\left. s_{k_1k_2k_3k_4}\left(\beta_0\right)\right| X_i, X_j, A_i, A_j, U_{ij}\right]. \tag{33}$$

The expectation in the summand of (33) is, by iterated expectations and conditionally independent link formation, zero if $\{i,j\} \nsubseteq \{k_1,k_2,k_3,k_4\}$ and equals $\bar{s}_{ij}\left(\beta_0\right)$ if $\{i,j\} \subseteq$

$\{k_1, k_2, k_3, k_4\}$. Since dyad $\{i, j\}$ appears in a total of $\binom{N-2}{2}$ of the $\binom{N}{4}$ tetrads we get that

$$\mathbb{E}\left[U_N | X_i, X_j, A_i, A_j, U_{ij}\right] = \binom{N}{4}^{-1} \binom{N-2}{2} = \frac{12}{N(N-1)} \bar{s}_{ij}(\beta_0) = \frac{6}{n} \bar{s}_{ij}(\beta_0).$$

Next observe that $\mathbb{E}\left[\bar{s}_{ij}(\beta_0) | X_k, X_l, A_k, A_l, U_{kl}\right] = 0$ for all $\{i, j\} \neq \{k, l\}$ (importantly this mean independence result holds true even if $\{i, j\}$ and $\{k, l\}$ share one index in common). In then follows that the Hajek Projection is (cf., van der Vaart, 2000, Lemma 11.10, p. 157)

$$U_N^* = \frac{6}{n} \sum_{i<j} \bar{s}_{ij}(\beta_0). \tag{34}$$

Asymptotic equivalence of $\sqrt{n\alpha_{2,N}^{-1}} U_N$ and $\sqrt{n\alpha_{2,N}^{-1}} U_N^*$ follows if

$$n\alpha_{2,N}^{-1} \mathbb{E}\left[(U_N^* - U_N)^2\right] = n\alpha_{2,N}^{-1} \mathbb{V}(U_N) + n\alpha_{2,N}^{-1} \mathbb{V}(U_N^*) - 2n\alpha_{2,N}^{-1} \mathbb{C}(U_N, U_N^*)$$

is $o_p(1)$. If $N\lambda_N \to \infty$, then the first term to the right of the equality in the expression above converges in probability to a constant since $\Delta_{2,N} = O(\alpha_{2,N})$. Next observe that while the the random variables $\{\bar{s}_{ij}(\beta_0)\}_{i<j}$ entering the sum in (34) are not independently and identically distributed, they are uncorrelated. This is an implication of conditionally independent edge formation given $\mathbf{X}$ and $\mathbf{A}$. This gives $\mathbb{V}(U_N^*) = \frac{36}{n} \Delta_{2,N}$ and hence that $n\alpha_{2,N}^{-1} \mathbb{V}(U_N^*)$ has the same probability limit as $n\alpha_{2,N}^{-1} \mathbb{V}(U_N)$. All that remains is to evaluate the covariance term:

$$\begin{aligned} n\alpha_{2,N}^{-1} \mathbb{C}(U_N, U_N^*) &= n\alpha_{2,N}^{-1} \mathbb{E}\left[U_N U_N^*\right] \\ &= n\alpha_{2,N}^{-1} \mathbb{E}\left[(U_N - U_N^*) U_N^*\right] + n\alpha_{2,N}^{-1} \mathbb{E}\left[U_N^* U_N^*\right] \\ &= \frac{36}{\alpha_{2,N}} \Delta_{2,N}, \end{aligned}$$

giving $n\alpha_{2,N}^{-1} \mathbb{E}\left[(U_N^* - U_N)^2\right] \to 0$ as $N \to \infty$.

Step 2c, CLT:

Putting the above results together we have that

$$\begin{aligned} \sqrt{n\alpha_{2,N}^{-1}} \alpha_{4,N} \left(\hat{\beta}_{\mathrm{TL}} - \beta_0\right) &= -\left[\alpha_{4,N}^{-1} \binom{N}{4}^{-1} \sum_{i<j<k<l} \frac{\partial^2 g_{ijkl}(\bar{\beta})}{\partial \beta \partial \beta'}\right]^{+} \times \sqrt{n\alpha_{2,N}^{-1}} U_N \\ &= -6\Gamma_0^{-1} \left[\frac{1}{\sqrt{n\alpha_{2,N}}} \sum_{i<j}^{N} \bar{s}_{ij}(\beta_0)\right] + o_p(1) \end{aligned} \tag{35}$$

The main result follows if we can demonstrate asymptotic normality of $\frac{1}{\sqrt{n\alpha_{2,N}}}\sum_{i<j}^{N}\bar{s}_{2,ij}$. This can be established by adapting an argument used by Chatterjee (2006) in the proof of his Theorem 1.1.

Recall that the boldface subscripts $\mathbf{i}=\mathbf{1},\mathbf{2},\ldots$ index the $n=\binom{N}{2}$ dyads in arbitrary order. Let $\tilde{\Delta}_{\mathbf{i}}=\alpha_{2,N}^{-1}\mathbb{E}\left[\bar{s}_{\mathbf{i}}\bar{s}_{\mathbf{i}}'\right|X_{i_1},X_{i_2},A_{i_1},A_{i_2}]$ denote the scaled conditional variance of the projection $\bar{s}_{\mathbf{i}}=\mathbb{E}\left[s_{ijkl}\right|X_i,X_j,A_i,A_j,U_{ij}]$. Observe that, by independence of U_{ij} and $(\mathbf{X},\mathbf{A})$ (Assumption 1) and random sampling (Assumption 3), we have that

$$\tilde{\Delta}_{\mathbf{i}}=\alpha_{2,N}^{-1}\mathbb{E}\left[\bar{s}_{\mathbf{i}}\bar{s}_{\mathbf{i}}'\right|X_{i_1},X_{i_2},A_{i_1},A_{i_2}]=\alpha_{2,N}^{-1}\mathbb{E}\left[\bar{s}_{\mathbf{i}}\bar{s}_{\mathbf{i}}'\right|\mathbf{X},\mathbf{A}].$$

Furthermore $\tilde{\Delta}_{\mathbf{i}}$ is bounded away from zero for any fixed N (by Assumptions 2 and 4.iii).

Next let c be a vector of real constants and define

$$R_{\mathbf{i}}=\frac{c'\Gamma_0^{-1}\left(\bar{s}_{\mathbf{i}}/\sqrt{\alpha_{2,N}}\right)}{\sqrt{c'\Gamma_0^{-1}\Delta_N\Gamma_0^{-1}c}},\tag{36}$$

where $\tilde{\Delta}_N=\frac{1}{n}\sum_{\mathbf{i}=1}^{n}\tilde{\Delta}_{\mathbf{i}}$. By the conditional mean zero property of the score function we get that $\mathbb{E}\left[R_{\mathbf{i}}\right|\mathbf{X},\mathbf{A}]=0$ and, by the arguments sketched above, that $\mathbb{E}\left[R_{\mathbf{i}}^2\right|\mathbf{X},\mathbf{A}]=\frac{c'\Gamma_0^{-1}\tilde{\Delta}_{\mathbf{i}}\Gamma_0^{-1}c}{c'\Gamma_0^{-1}\tilde{\Delta}_N\Gamma_0^{-1}c}$ (observe that $\tilde{\Delta}_N$ is a function of $\mathbf{X}$ and $\mathbf{A}$). Let $\mathbf{Y}$ be a $n\times 1$ random vector with independent non-identically distributed normal components $Y_{\mathbf{i}}\sim\mathcal{N}\left(0,\frac{c'\Gamma_0^{-1}\tilde{\Delta}_{\mathbf{i}}\Gamma_0^{-1}c}{c'\Gamma_0^{-1}\tilde{\Delta}_N\Gamma_0^{-1}c}\right)$. Let $\mathcal{C}_M$ denote the class of functions $f:\mathbb{R}\to\mathbb{R}$ that are three times continuously differentiable with $\sup_x\left|\frac{\partial^r f(x)}{\partial x^r}\right|<L_r(f)<\infty$ for $r=1,2,3$. The proof proceeds by showing that $\mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}R_{\mathbf{i}}\right)\right]\to\mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}Y_{\mathbf{i}}\right)\right]$ as $N\to\infty$ for each f in the class $\mathcal{C}_M$. Since $\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}Y_{\mathbf{i}}\sim\mathcal{N}(0,1)$ by construction, this implies that $\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n}R_{\mathbf{i}}\xrightarrow{D}\mathcal{N}(0,1)$ (e.g., Pollard, 2002, Lemma 16, p. 177).

To show this result I begin, as in Chatterjee (2006, p. 2065), by defining

$$\mathbf{Z}_{\mathbf{i}}=(R_{\mathbf{1}},\ldots,R_{\mathbf{i}},Y_{\mathbf{i}+1},\ldots,Y_n)\text{ and }\mathbf{Z}_{\mathbf{i}}^0=(R_{\mathbf{1}},\ldots,R_{\mathbf{i}-1},0,Y_{\mathbf{i}+1},\ldots,Y_n),$$

and observe that we can re-write the difference $f\left(\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n} R_{\mathbf{i}}\right) - f\left(\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n} Y_{\mathbf{i}}\right)$ as

$$\begin{aligned} f\left(\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n} R_{\mathbf{i}}\right) - f\left(\frac{1}{\sqrt{n}}\sum_{\mathbf{i}=1}^{n} Y_{\mathbf{i}}\right) &= \sum_{\mathbf{i}=1}^{n} f\left(\frac{1}{\sqrt{n}}\left[\sum_{\mathbf{j}=1}^{\mathbf{i}} R_{\mathbf{j}} + \sum_{\mathbf{j}=\mathbf{i}+1}^{\mathbf{n}} Y_{\mathbf{j}}\right]\right) \\ &\quad - f\left(\frac{1}{\sqrt{n}}\left[\sum_{\mathbf{j}=1}^{\mathbf{i}-1} R_{\mathbf{j}} + \sum_{\mathbf{j}=\mathbf{i}}^{\mathbf{n}} Y_{\mathbf{j}}\right]\right) \\ &= \sum_{\mathbf{i}=1}^{n} f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i}}'\iota\right) - f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i-1}}'\iota\right). \end{aligned} \tag{37}$$

A third-order Taylor expansion of $f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i}}'\iota\right)$ about $R_{\mathbf{i}} = 0$ gives

$$\begin{aligned} \Bigg| f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i}}'\iota\right) - f\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right) - \frac{1}{\sqrt{n}}f'\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right) R_{\mathbf{i}} \\ -\frac{1}{2}\frac{1}{n}f''\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right) R_{\mathbf{i}}^{2}\Bigg| \leq \frac{\left|R_{\mathbf{i}}^{3}\right| L_3(f)}{6n^{3/2}} \end{aligned} \tag{38}$$

for $\mathbf{i} = 1, \ldots, n$. A second expansion of $f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i}-1}'\iota\right)$ about $Y_{\mathbf{i}} = 0$ similarly gives

$$\begin{aligned} \Bigg| f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i}-1}'\iota\right) - f\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right) - \frac{1}{\sqrt{n}}f'\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right) Y_{\mathbf{i}} \\ -\frac{1}{2}\frac{1}{n}f''\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right) Y_{\mathbf{i}}^{2}\Bigg| \leq \frac{\left|Y_{\mathbf{i}}^{3}\right| L_3(f)}{6n^{3/2}} \end{aligned} \tag{39}$$

also for for $\mathbf{i} = 1, \ldots, n$.

Since $R_{\mathbf{i}}$ is independent of $R_{\mathbf{j}}$ for $\mathbf{i} \neq \mathbf{j}$ conditional on $\mathbf{X}$ and $\mathbf{A}$ and the the $Y_{\mathbf{i}}'s$ are independent by construction we have that

$$\mathbb{E}\left[\left(R_{\mathbf{i}} - Y_{\mathbf{i}}\right)\frac{1}{\sqrt{n}}f'\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right)\right] = 0. \tag{40}$$

Similarly, we have that

$$\mathbb{E}\left[\left(R_{\mathbf{i}}^{2} - Y_{\mathbf{i}}^{2}\right)\frac{1}{n}f''\left(\frac{1}{\sqrt{n}}\left(\mathbf{Z}_{\mathbf{i}}^{0}\right)'\iota\right)\right] = 0 \tag{41}$$

Equations (38), (39), (40) and (41) imply

$$\left|\mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i}}'\iota\right)\right] - \mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}\mathbf{Z}_{\mathbf{i-1}}'\iota\right)\right]\right| \leq \frac{\mathbb{E}\left[\left|R_{\mathbf{i}}^{3}\right| + \left|Y_{\mathbf{i}}^{3}\right|\right]}{6n^{3/2}} L_3(f)$$

Summing over $\mathbf{i} = 1, \ldots, n$ and using (37) we therefore have

$$\left| \mathbb{E}\left[f\left(\frac{1}{\sqrt{n}} \sum_{\mathbf{i}=1}^{n} R_{\mathbf{i}} \right) \right] - \mathbb{E}\left[f\left(Z \right) \right] \right| \leq \frac{1}{6} \frac{M_3}{\sqrt{n}} L_3\left(f \right)$$

with $M_3 = \max_i \mathbb{E}\left[|R_{\mathbf{i}}|^3 \right]$ (which is bounded by Assumption 2 and 4.iii; note also that $\mathbb{E}\left[|Y_{\mathbf{i}}|^3 \right] = 0$ by normality) and Z a standard normal random variable.

Using this result, (35) and (36) I get

$$\begin{aligned} \frac{\alpha_{4,N} \sqrt{n \alpha_{2,N}^{-1}} c' \left(\hat{\beta}_{\mathrm{TL}} - \beta_0 \right)}{\sqrt{c' \Gamma_0^{-1} \tilde{\Delta}_N \Gamma_0^{-1} c}} &= 6 \left[\frac{1}{\sqrt{n}} \sum_{i<j}^{N} \frac{c' \Gamma_0^{-1} \left(\bar{s}_{ij} / \sqrt{\alpha_{2,N}} \right)}{\sqrt{c' \Gamma_0^{-1} \tilde{\Delta}_N \Gamma_0^{-1} c}} \right] + o_p\left(1 \right) \\ &= 6 \left[\frac{1}{\sqrt{n}} \sum_{\mathbf{i}=1}^{n} R_{\mathbf{i}} \right] + o_p\left(1 \right) \\ &\xrightarrow{D} \mathcal{N}\left(0, 36 \right) \end{aligned}$$

as claimed.

References

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics*, 74(1):47 – 97.

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

Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26(1):31 – 44.

Apicella, C. L., Marlowe, F. W., Fowler, J. H., and Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers,. *Nature*, 481(7382):497 – 501.

Arellano, M. and Hahn, J. (2007). *Advances in Economics and Econometrics: Theory and Applications*, volume 3, chapter Understanding bias in nonlinear panel models: some recent developments, pages 381 – 309. Cambridge University Press., Cambridge.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., and Meghir, C. (2012). Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, 4(2):134 – 167.

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, 68(5):1181 – 1229.

Banerjee, A., Chandrasekhar, A. G., Dulfo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):363 – 370.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509 – 512.

Bhattacharya, S. and Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics*, 43(6):2384 – 2411.

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, Inc., New York, 3rd edition.

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

Bloch, F. and Jackson, M. O. (2007). The formation of networks with transfers among players. *Journal of Economic Theory*, 113(1):83 – 110.

Candelaria, L. E. (2016). A semiparametric network formation model with multiple linear fixed effects. Duke University.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225 – 238.

Chamberlain, G. (1984). *Handbook of Econometrics*, volume 2, chapter Panel Data, pages 1247 – 1318. North-Holland, Amsterdam.

Chamberlain, G. (1985). *Longitudinal Analysis of Labor Market Data*, chapter Heterogeneity, omitted variable bias, and duration dependence, pages 3 – 38. Cambridge University Press, Cambridge.

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

Charbonneau, K. B. (2014). Multiple fixed effects in binary response panel data models. Technical Report Working Paper 2014-17, Bank of Canada.

Chatterjee, S. (2006). A generalization of the lindeberg principle. *Annals of Probability*, 34(6):2061 – 2076.

Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability*, 21(4):1400 – 1435.

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

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society B*, 20(2):215 – 242.

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

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

De Weerdt, J. (2004). *Insurance Against Poverty*, chapter Risk-sharing and endogenous network formation, pages 197 – 216. Oxford University Press, Oxford.

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

Fafchamps, M. and Gubert, F. (2007). Risk sharing and network formation. *American Economic Review*, 97(2):75 – 79.

Faust, K. (2007). Very local structure in social networks. *Sociological Methodology*, 37(1):209 – 256.

Ferguson, T. S. (2005). U-statistics. University of California - Los Angeles.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel data models with large n, t. *Journal of Econometrics*, 192(1):291 – 312.

Fox, J. T. (2010). Identification in matching games. *Quantitative Economics*, 1(2):203 – 254.

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.

Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics*, 127(3):1287 – 1338.

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

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. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7(1):465–485.

Graham, B. S. (2016). Homophily and transitivity in dynamic network formation. NBER Working Paper 22186, National Bureau of Economic Research.

Hahn, J., Moon, H. R., and Snider, C. (2016). Lm test of neglected correlated random effects and its application. University of California - Los Angeles.

Hahn, J. and Newey, W. K. (2004). Jackknife and analytical bias reduction for nonlinear panel data models. *Econometrica*, 72(4):1295 – 1319.

Hillar, C. J., Lin, S., and Wibisono, A. (2013). Inverses of symmetric, diagonally dominant positive matrices and applications. Technical Report arXiv:1203.6812v2, University of California - Berkeley.

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

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13 – 30.

Holland, P. W. and Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7:1 – 45.

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 Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, 64(1-2):241 – 278.

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*. Cambridge University Press, Cambridge, 2nd edition.

Jackson, M. O. and López-Pintado, D. (2013). Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, 1(1):49 – 67.

Jackson, M. O. and Rogers, B. W. (2007). Relating network structure to diffusion properties through stochastic dominance. *B.E. Journal of Theoretical Economics*, 7(1):(Advances) Article 6.

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.

Jochmans, K. (2016a). Modified-likelihood estimation of fixed-effect models for dyadic data. Technical Report Working Paper 1, Sciences Po. Sciences Po.

Jochmans, K. (2016b). Semiparametric analysis of network formation. Technical Report Working Paper 2, Sciences Po.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(2):204 – 213.

Lai, B. and Reiter, D. (2000). Democracy, political similarity, and international alliances, 1816-1992. *Journal of Conflict Resolution*, 44(2):203 – 227.

Leung, M. (2015). A random-field approach to inference in large models of network formation. Mimeo, Stanford University.

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

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27(1):415 – 444.

Mele, A. (2013). A structural model of segregation in social networks. John Hopkins University.

Menzel, K. (2016). Strategic network formation with many agents. New York University.

Miyauchi, Y. (2016). Structural estimation of a pairwise stable network with nonnegative externality. *Journal of Econometrics*, 195(2):224 – 235.

Nadler, C. (2015). Networked inequality: evidence from freelancers. University of California - Berkeley.

Newey, W. K. and McFadden, D. (1994). *Handbook of Econometrics*, volume 4, chapter Large sample estimation and hypothesis testing, pages 2111 – 2245. North-Holland, Amsterdam.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press, Oxford.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1 – 32.

Olhede, S. C. and Wolfe, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 11(41):14722 – 14727.

Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200 – 3203.

Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge.

Rao, B. L. S. P. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics*, 61(2):441 – 460.

Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4):641 – 658.

Shalizi, C. R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *Annals of Statistics*, 41(2):508 – 535.

Sheng, S. (2014). A structural econometric analysis of network formation games. Mimeo, University of California - Los Angeles.

Simons, G. and Yao, Y.-C. (1998). Approximating the inverse of a symmetric positive definite matrix. *Linear Algebra and its Applications*, 281(1-3):97 – 103.

Simons, G. and Yao, Y.-C. (1999). Asymptotics when the number of parameters tends to infinity in the bradley-terry model for paired comparisons. *Annals of Statistics*, 27(3):1041 – 1060.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies*, 70(1):147 – 165.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

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.

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. (2016). Statistical inference in a directed network model with covariates. arxiv 1609.04558, University of Warwick.

Yan, T. and Xu, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika*, 100(2):519 – 524.